

# On the limiting law of the length of the longest common and increasing subsequences in random words\*

Jean-Christophe Breton<sup>†</sup>

IRMAR, UMR 6625, Université de Rennes 1,  
263 Avenue du Général Leclerc CS 74205, 35042, Rennes, France

Christian Houdré<sup>‡</sup>

School of Mathematics, Georgia Institute of Technology,  
Atlanta, GA 30332, USA

À la Mémoire de Marc Yor

## Abstract

Let  $X = (X_i)_{i \geq 1}$  and  $Y = (Y_i)_{i \geq 1}$  be two sequences of independent and identically distributed (iid) random variables taking their values, uniformly, in a common totally ordered finite alphabet. Let  $\text{LCI}_n$  be the length of the longest common and (weakly) increasing subsequence of  $X_1 \cdots X_n$  and  $Y_1 \cdots Y_n$ . As  $n$  grows without bound, and when properly centered and scaled,  $\text{LCI}_n$  is shown to converge, in distribution, towards a Brownian functional that we identify.

**Keywords:** Longest Common Subsequence, Longest Increasing Subsequence, Random Words, Random Matrices, Donsker's Theorem, Optimal Alignment, Last Passage Percolation.

**Mathematics Subject Classification (2010):** 05A05, 60C05, 60F05.

## 1 Introduction

We analyze below the asymptotic behavior of the length of the longest common subsequence in random words with an additional (weakly) increasing requirement. Although it has been studied from an algorithmic point of view in computer science, bio-informatics, or statistical physics (see, for instance, [CZFYZ], [DKFPWS] or [Sak]), to name but a few fields, mathematical results for this hybrid problem are very sparse. To present our

---

\*Final version to appear in Stochastic Processes and Their Applications

<sup>†</sup>E-mail: [jean-christophe.breton@univ-rennes1.fr](mailto:jean-christophe.breton@univ-rennes1.fr)

<sup>‡</sup>E-mail: [houdre@math.gatech.edu](mailto:houdre@math.gatech.edu)

framework, let  $X = (X_i)_{i \geq 1}$  and  $Y = (Y_i)_{i \geq 1}$  be two infinite sequences whose coordinates take their values in  $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < \dots < \alpha_m\}$ , a finite totally ordered alphabet of cardinality  $m$ . Next,  $\text{LCI}_n$ , the length of the longest common and (weakly) increasing subsequences of the words  $X_1 \dots X_n$  and  $Y_1 \dots Y_n$  is the maximal integer  $k \in \{1, \dots, n\}$ , such that there exist  $1 \leq i_1 < \dots < i_k \leq n$  and  $1 \leq j_1 < \dots < j_k \leq n$ , satisfying the following two conditions:

- (i)  $X_{i_s} = Y_{j_s}$ , for all  $s = 1, 2, \dots, k$ ,
- (ii)  $X_{i_1} \leq X_{i_2} \leq \dots \leq X_{i_k}$  and  $Y_{j_1} \leq Y_{j_2} \leq \dots \leq Y_{j_k}$ .

(Asymptotically, the strictly increasing case is of little interest, having  $m$  as a pointwise limiting behavior.)  $\text{LCI}_n$  is a measure of the similarity/dissimilarity of the random words often used in pattern matching, and its asymptotic behavior is the purpose of our study. This limiting behavior differs from the one of another better-known, measure of similarity/dissimilarity, namely,  $\text{LC}_n$ , the length of the longest common subsequences of two or more random words. Indeed, after renormalization, the first result on  $\text{LC}_n$ , obtained in [HI], reveals, under a sublinear variance lower bound assumption, a normal limiting law. In contrast, for  $\text{LCI}_n$ , we have:

**Theorem 1.1** *Let  $X = (X_i)_{i \geq 1}$  and  $Y = (Y_i)_{i \geq 1}$  be two sequences of iid random variables uniformly distributed on  $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < \dots < \alpha_m\}$ , a totally ordered finite alphabet of cardinality  $m$ . Let  $\text{LCI}_n$  be the length of the longest common and increasing subsequences of  $X_1 \dots X_n$  and  $Y_1 \dots Y_n$ . Then, as  $n \rightarrow +\infty$ ,*

$$\frac{\text{LCI}_n - n/m}{\sqrt{n/m}} \Rightarrow \max_{0=t_0 \leq t_1 \leq \dots \leq t_{m-1} \leq t_m=1} \min \left( -\frac{1}{m} \sum_{i=1}^m B_1^{(i)}(1) + \sum_{i=1}^m \left( B_1^{(i)}(t_i) - B_1^{(i)}(t_{i-1}) \right), \right. \\ \left. -\frac{1}{m} \sum_{i=1}^m B_2^{(i)}(1) + \sum_{i=1}^m \left( B_2^{(i)}(t_i) - B_2^{(i)}(t_{i-1}) \right) \right) \quad (1.1)$$

where  $B_1$  and  $B_2$  are two  $m$ -dimensional standard Brownian motions on  $[0, 1]$ .

The main motivation for our work has its origins in the identification, first obtained by Kerov [Ker], of the limiting length (properly centered and scaled) of the longest increasing subsequence of a random word, as the maximal eigenvalue of a certain Gaussian random matrix. When combined with results of Baryshnikov [Bar] or Gravner, Tracy and Widom [GTW] (see also [BGH]), this limiting law has a representation as a Brownian functional. Moreover, the longest increasing subsequence corresponds to the first row of the RSK Young diagrams associated with the random word and [Ker, Chap. 3, Sec. 3.4, Theorem 2] showed that the whole normalized limiting shape of these RSK Young diagrams is the spectrum of the traceless Gaussian Unitary Ensemble (GUE). Since the length of the top row of the diagrams is the length of the longest increasing subsequence of the random word, the maximal eigenvalue result is recovered. (The asymptotic length

result was rediscovered by Tracy and Widom [TW] and the asymptotic shape one by Johansson [Joh]. Extensions to non-uniform letters were also obtained by Its, Tracy and Widom [ITW1, ITW2].) Another motivation for the present study comes from the interpretation of the  $\text{LCI}_n$  functional in terms of last passage time in directed percolation. This is detailed in our concluding remarks.

The asymptotic behavior of the length of the longest common and increasing subsequences has actually already been investigated for binary words ( $m = 2$ ) in [HLM]. However, the methods used there do not allow to consider an alphabet of arbitrary finite size  $m$ . When  $m = 2$  with letters  $\alpha_1$  and  $\alpha_2$ , it is enough to consider common subsequences made of a random number of common  $\alpha_1$ 's deterministically completed by the common  $\alpha_2$ 's, so that in a way the corresponding study is reduced to deal with only one type of letter. In contrast, when  $m \geq 3$ , the situation is much more complicated since a similar strategy reduced the problem to  $m - 1$  types of letter for which there is still, roughly speaking, too much randomness to successfully handle, in this way, the study of  $\text{LCI}_n$ . A new methodology based on a new representation of  $\text{LCI}_n$  is thus required to deal with general finite alphabet of size  $m$ . This is achieved below where an appropriate representation of  $\text{LCI}_n$ , that allows to investigate its asymptotic behavior for arbitrary  $m \geq 2$ , is obtained. Our results thus extend and encompass the binary  $\text{LCI}_n$  result of [HLM]. The dependence (or independence) structure between the two sequences of letters  $X$  and  $Y$  is carried over at the limit into a similar structure between the two standard Brownian motions  $B_1$  and  $B_2$ . Hence, when  $X = Y$ , our results recover, with the help of [BGH], the weak limits obtained in [Ker], [Joh], [TW], [ITW1], [ITW2], [HL], and [HX], while if  $X$  and  $Y$  are independent so are  $B_1$  and  $B_2$ . As a by-product of our approach, we further fix some loose points present in [HLM]. As suggested to us, let us further put our main theorem in context. At first, for  $m = 2$ , the right hand-side of (1.1) becomes

$$\max_{0 \leq t \leq 1} \min \left( \frac{B_1^{(2)}(1) - B_1^{(1)}(1)}{2} - (B_1^{(2)}(t) - B_1^{(1)}(t)), \frac{B_2^{(2)}(1) - B_2^{(1)}(1)}{2} - (B_2^{(2)}(t) - B_2^{(1)}(t)) \right).$$

In case the two-dimensional standard Brownian motions are independent, this last expression has the same law as

$$\sqrt{2} \max_{0 \leq t \leq 1} \min \left( B_1(t) - \frac{1}{2}B_1(1), B_2(t) - \frac{1}{2}B_2(1) \right),$$

where, now,  $B_1$  and  $B_2$  are two independent one-dimensional standard Brownian motions on  $[0, 1]$ . Therefore, our limiting result matches the binary one presented in [HLM].

Next, and still for further context, let us compare the asymptotic behavior of  $\text{LCI}_n$  to the one of, say,  $L_n$ , the length of the optimal alignments which align only one type of letters. (In case of a single word,  $L_n$  could correspond to, e.g, the length of the longest constant subsequences). Clearly,  $\text{LCI}_n \geq L_n$  and under a uniform assumption,

$$\lim_{n \rightarrow +\infty} \frac{\text{LCI}_n}{n} = \lim_{n \rightarrow +\infty} \frac{L_n}{n} = \frac{1}{m}, \quad (1.2)$$

with probability one. Moreover, it is easy to see that, as  $n \rightarrow +\infty$ ,

$$\frac{L_n - n/m}{\sqrt{n/m}} \Rightarrow \min \left( \sqrt{1 - \frac{1}{m}} B_1(1), \sqrt{1 - \frac{1}{m}} B_2(1) \right), \quad (1.3)$$

for, say, two one-dimensional standard Brownian motions  $B_1$  and  $B_2$ . Now returning to (1.1), note that for  $j = 1, 2$ ,

$$\begin{aligned} & -\frac{1}{m} \sum_{i=1}^m B_j^{(i)}(1) + \sum_{i=1}^m (B_j^{(i)}(t_i) - B_j^{(i)}(t_{i-1})) \\ &= \frac{1}{m} \left( (m-1) B_j^m(1) - \sum_{i=1}^{m-1} B_j^{(i)}(1) \right) + \sum_{i=1}^{m-1} (B_j^{(i)}(t_i) - B_j^{(i)}(t_{i-1})) - B_j^{(m)}(t_{m-1}), \end{aligned} \quad (1.4)$$

where the random variable  $((m-1) B_j^m(1) - \sum_{i=1}^{m-1} B_j^{(i)}(1))/m$  has exactly the same law as  $\sqrt{1 - 1/m} B_j(1)$ . Therefore, the presence of the extra terms involving the  $t'_i$ s on the right hand-side of (1.1) allows to distinguish the renormalized limit of  $\text{LCI}_n$  from that of  $L_n$  and ensures that the latter limit is still almost surely dominated by the former. This observation should be contrasted with the non-uniform case where a single letter is attained with maximal probability  $p_{\max}$ , and where  $L_n$  aligns this letter. Indeed, in view of (5.1), below, when centered by  $np_{\max}$  and scaled by  $\sqrt{np_{\max}}$ , both  $\text{LCI}_n$  and  $L_n$  converge to  $\min(\sqrt{1 - p_{\max}} B_1(1), \sqrt{1 - p_{\max}} B_2(1))$ .

A natural question arising from this study is the random matrix interpretation of our limiting ditribution (1.1). Another natural question is to interpret  $\text{LCI}_n$  in terms of RSK Young diagrams and to investigate, more generally, the shape of a RSK counterpart of  $\text{LCI}_n$ . Both questions go actually far beyond the scope of this paper but will be the subject of forthcoming investigations.

As for the content of the paper, the next section (Section 2) establishes a pathwise representation for the length of the longest common and increasing subsequence of the two words as a max/min functional. In Section 3, the probabilistic framework is initiated, the representation becomes the maximum over a random set of the minimum of random sums of randomly stopped random variables. The various random variables involved are studied and their (conditional) laws found. In Section 4, the limiting law is obtained. This is done in part by a derandomization procedure (of the random sums and of the random constraints) leading to the Brownian functional (1.1) of Theorem 1.1. In the last section (Section 5), various extensions and generalizations are discussed as well as some open questions related to this problem. Finally, Appendix A.1 completes the proof of some technical results and while Appendix A.2 gives missing steps in the proof of the main theorem in [HLM] as well as corrections to arguments presented there; providing, in the much simpler binary case, a rather self-contained proof.

## 2 Combinatorics

The aim of this section is to obtain a pathwise representation for the length of the longest common and increasing subsequences of two finite words. Throughout the paper,  $X = (X_i)_{i \geq 1}$  and  $Y = (Y_i)_{i \geq 1}$  are two infinite sequences whose coordinates take their values in  $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < \dots < \alpha_m\}$ , a finite totally ordered alphabet of cardinality  $m$ . Recall next that  $\text{LCI}_n$  is the maximal integer  $k \in \{1, \dots, n\}$ , such that there exist  $1 \leq i_1 < \dots < i_k \leq n$  and  $1 \leq j_1 < \dots < j_k \leq n$ , satisfying the following two conditions:

- (i)  $X_{i_s} = Y_{j_s}$ , for all  $s = 1, 2, \dots, k$ ,
- (ii)  $X_{i_1} \leq X_{i_2} \leq \dots \leq X_{i_k}$  and  $Y_{j_1} \leq Y_{j_2} \leq \dots \leq Y_{j_k}$ .

Now that  $\text{LCI}_n$  has been formally defined, let us set some standing notation. Let  $N_r(X)$ ,  $r = 1, \dots, m$ , be the number of  $\alpha_r$ s in  $X_1, X_2, \dots, X_n$ , i.e.,

$$N_r(X) = \#\{i = 1, \dots, n : X_i = \alpha_r\} = \sum_{i=1}^n \mathbf{1}_{\{X_i = \alpha_r\}}, \quad (2.1)$$

and similarly let  $N_r(Y)$ ,  $r = 1, \dots, m$ , be the number of  $\alpha_r$ s in  $Y_1, Y_2, \dots, Y_n$ . Clearly,

$$\sum_{r=1}^m N_r(X) = \sum_{r=1}^m N_r(Y) = n.$$

Let us further set a convention: *Throughout the paper when there is no ambiguity or when a property is valid for both sequences  $X = (X_i)_{i \geq 1}$  and  $Y = (Y_i)_{i \geq 1}$  we often omit the symbol  $X$  or  $Y$  and, e.g., write  $N_r$  for either  $N_r(X)$  or  $N_r(Y)$  or, below,  $H$  for either  $H_X$  or  $H_Y$ .*

Continuing on our notational path, for each  $r = 1, \dots, m$ , let  $N_r^{s,t}(X)$  be the number of  $\alpha_r$ s in  $X_{s+1}, X_{s+2}, \dots, X_t$ , i.e.,

$$N_r^{s,t}(X) = \#\{i = s+1, \dots, t : X_i = \alpha_r\} = \sum_{i=s+1}^t \mathbf{1}_{\{X_i = \alpha_r\}}, \quad (2.2)$$

with a similar definition for  $N_r^{s,t}(Y)$ . Again, it is trivially verified that

$$\sum_{r=1}^m N_r^{s,t}(X) = \sum_{r=1}^m N_r^{s,t}(Y) = t - s,$$

and, of course,  $N_r^{0,n} = N_r$ . Still continuing with our notations, let  $T_r^j(X)$ ,  $r = 1, \dots, m$ , be the location of the  $j^{\text{th}}$   $\alpha_r$  in the *infinite* sequence  $X_1, X_2, \dots, X_n, \dots$ , with the convention that  $T_r^0(X) = 0$ . Then, for  $j = 1, 2, \dots$ ,  $T_r^j(X)$  can be defined recursively via,

$$T_r^j(X) = \min \{s \in \mathbb{N} : s > T_r^{j-1}(X), X_s = \alpha_r\} \quad (2.3)$$

where as usual  $\mathbb{N} = \{0, 1, 2, \dots\}$ . Again replacing  $X$  by  $Y$  gives the corresponding notion for the sequence  $Y = (Y_i)_{i \geq 1}$ .

Next, let us begin our finding of a representation for  $\text{LCI}_n$  via the random variables defined to date. First, let  $H_X(k_1, k_2, \dots, k_{m-1})$  be the maximal number of  $\alpha_m$ s contained in an increasing subsequence, of  $X_1 X_2 \dots X_n$ , containing  $k_1$   $\alpha_1$ s,  $k_2$   $\alpha_2$ s,  $\dots$ ,  $k_{m-1}$   $\alpha_{m-1}$ s picked in that order. Replacing  $X = (X_i)_{i \geq 1}$  by  $Y = (Y_i)_{i \geq 1}$ , it is then clear that

$$\min \left( k_1 + \dots + k_{m-1} + H_X(k_1, \dots, k_{m-1}), k_1 + \dots + k_{m-1} + H_Y(k_1, \dots, k_{m-1}) \right), \quad (2.4)$$

is, therefore, the length of the longest common and increasing subsequence of  $X_1 X_2 \dots X_n$  and  $Y_1 Y_2 \dots Y_n$  containing exactly  $k_r$   $\alpha_r$ s, for all  $r = 1, 2, \dots, m-1$ , the letters being picked in an increasing order. Hence, to find  $\text{LCI}_n$ , the function  $H$  needs to be identified and (2.4) needs to be maximized over all possible choices of  $k_1, k_2, \dots, k_{m-1}$ .

Let us start with the maximizing constraints. Assume, for a while, that a single word, say,  $X_1 \dots X_n$ , is considered. First, and clearly,  $0 \leq k_1 \leq N_1$ . Next,  $k_2$  is the number of  $\alpha_2$ s present in the sequence after the  $k_1^{\text{th}}$   $\alpha_1$ . Any letter  $\alpha_2$  is admissible but the ones occurring before the  $k_1^{\text{th}}$   $\alpha_1$ , attained at the location  $T_1^{k_1} \wedge n$ . Since there are  $n$  letters, considered so far, there are thus  $N_2^{0, T_1^{k_1} \wedge n}$  inadmissible  $\alpha_2$ s and the requirement on  $k_2$  writes  $k_2 \leq N_2 - N_2^{0, T_1^{k_1} \wedge n}$ . Similarly for each  $r = 3, \dots, m-1$ ,  $k_r$  is the number of letters  $\alpha_r$  minus the inadmissible  $\alpha_r$ s which occur during the recuperation, of the  $k_1$   $\alpha_1$ s, followed by the  $k_2$   $\alpha_2$ s, followed by the  $k_3$   $\alpha_3$ s, *etc* in that order. Thus the requirement on  $k_r$  is of the form  $k_r \leq N_r - \tilde{N}_r^*$ , where  $\tilde{N}_r^*$  is the number of  $\alpha_r$ s occurring before the  $k_i$   $\alpha_i$ s,  $i \leq r-1$ , picked in the order just described. For  $r = 1, 2$ , and as already shown,  $\tilde{N}_1^* = 0$  and  $\tilde{N}_2^* = N_2^{0, T_1^{k_1} \wedge n}$ . Assume next that, for  $r \geq 3$ ,  $\tilde{N}_{r-1}^*$  is well defined, then  $\tilde{N}_r^*$  is the number of  $\alpha_r$ s occurring before, in that order, the  $k_1$   $\alpha_1$ s,  $\dots$ , the  $k_{r-1}$   $\alpha_{r-1}$ s. A little moment of reflection makes it clear that the location of the  $k_{r-1}^{\text{th}}$  such  $\alpha_{r-1}$  is  $T_{r-1}^{k_{r-1} + \tilde{N}_{r-1}^*}$ , from which it recursively follows that:

$$\tilde{N}_r^* = N_r^{0, T_{r-1}^{k_{r-1} + \tilde{N}_{r-1}^*} \wedge n}.$$

**Remark 2.1** Note that  $\tilde{N}_r^*$  as well as  $N_r^*$  defined below in (2.8) actually depend on  $k_1, \dots, k_{r-1}$ , but in order to not overload our notation we will omit this dependency thereafter.

Returning to two sequences  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_n$ , the condition on  $k_r$ ,  $1 \leq r \leq m-1$ , writes as

$$0 \leq k_r \leq \left( N_r(X) - \tilde{N}_r^*(X) \right) \wedge \left( N_r(Y) - \tilde{N}_r^*(Y) \right).$$

From these choices of indices and (2.4),

$$\text{LCI}_n = \max_{\vec{c}_n} \min \left( \sum_{i=1}^{m-1} k_i + H_X(k_1, \dots, k_{m-1}), \sum_{i=1}^{m-1} k_i + H_Y(k_1, \dots, k_{m-1}) \right), \quad (2.5)$$

where the outer maximum is taken over  $(k_1, \dots, k_{m-1})$  in

$$\tilde{\mathcal{C}}_n = \left\{ (k_1, \dots, k_{m-1}) : k_1 \in \tilde{\mathcal{C}}_{n,1}, k_2 \in \tilde{\mathcal{C}}_{n,2}(k_1), k_3 \in \tilde{\mathcal{C}}_{n,3}(k_1, k_2), k_{m-1} \in \tilde{\mathcal{C}}_{n,m-1}(k_1, \dots, k_{m-2}) \right\}, \quad (2.6)$$

where  $\tilde{\mathcal{C}}_{n,1} = \left\{ 0 \leq k_1 \leq (N_1(X) - \tilde{N}_1^*(X)) \wedge (N_1(Y) - \tilde{N}_1^*(Y)) \right\}$  and for  $i = 2, \dots, m-1$ ,

$$\tilde{\mathcal{C}}_{n,i}(k_1, \dots, k_{i-1}) = \left\{ 0 \leq k_i \leq (N_i(X) - \tilde{N}_i^*(X)) \wedge (N_i(Y) - \tilde{N}_i^*(Y)) \right\}. \quad (2.7)$$

Next, observe that if  $T_{r-1}^{k_{r-1} + \tilde{N}_{r-1}^*} > n$ , then  $N_r - \tilde{N}_r^* = 0$ . Also, since the above maximum does not change under vacuous constraints, one can replace in the defining constraints,  $\tilde{N}_r^*$  by  $N_r^*$  recursively given via:  $N_1^* = 0$  and for  $r = 2, \dots, m-1$ ,

$$N_r^* = N_r^{0, T_{r-1}^{k_{r-1} + N_{r-1}^*}}. \quad (2.8)$$

The combinatorial expression (2.5) then becomes

$$\text{LCI}_n = \max_{\mathcal{C}_n} \min \left( \sum_{i=1}^{m-1} k_i + H_X(k_1, \dots, k_{m-1}), \sum_{i=1}^{m-1} k_i + H_Y(k_1, \dots, k_{m-1}) \right),$$

where the outer maximum is taken over  $(k_1, \dots, k_{m-1})$  in  $\mathcal{C}_n$  with  $\mathcal{C}_n$  and  $\mathcal{C}_{n,i}$ ,  $i = 1, \dots, m-1$ , respectively defined as in (2.6) and in (2.7) but with  $\tilde{N}_i^*$  replaced by  $N_i^*$ ,  $i = 1, \dots, m-1$ . and, of course,

$$\sum_{i=1}^m N_i(X) = \sum_{i=1}^m N_i(Y) = n.$$

After this identification, recall that  $H$  is the maximal number of  $\alpha_m$  after, in that order, the  $k_1$   $\alpha_1$ s,  $k_2$   $\alpha_2$ s,  $\dots$ ,  $k_{m-1}$   $\alpha_{m-1}$ s. Counting the  $\alpha_m$ s present between the various locations of the  $\alpha_i$ ,  $i = 1, \dots, m-1$ , and after another moment of reflection, it is clear that

$$H = N_m - R,$$

where

$$R = \sum_{i=1}^{m-1} \sum_{j=N_i^*+1}^{N_i^*+k_i} N_m^{T_i^{j-1}, T_i^j}, \quad (2.9)$$

and where the  $N_i^*$  are given by (2.8). Recall also that according to Remark 2.1,  $R$  depends actually on  $k_1, \dots, k_{m-1}$  but that for the sake of readability this dependency is omitted from our notations. Summarizing our results leads so far to:

**Theorem 2.1** *Let  $X = (X_i)_{i \geq 1}$  and  $Y = (Y_i)_{i \geq 1}$  be two sequences whose coordinates take their values in  $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < \dots < \alpha_m\}$ , a totally ordered finite alphabet of*

cardinality  $m$ . Let  $LCI_n$  be the length of the longest common and increasing subsequences of  $X_1 \cdots X_n$  and  $Y_1 \cdots Y_n$ . Then,

$$LCI_n = \max_{\mathcal{C}_n} \min \left( \sum_{i=1}^{m-1} k_i + N_m(X) - R(X), \sum_{i=1}^{m-1} k_i + N_m(Y) - R(Y) \right), \quad (2.10)$$

where the outer maximum is taken over  $(k_1, \dots, k_{m-1})$  in

$$\mathcal{C}_n = \left\{ (k_1, \dots, k_{m-1}) : k_1 \in \mathcal{C}_{n,1}, k_2 \in \mathcal{C}_{n,2}(k_1), k_3 \in \mathcal{C}_{n,3}(k_1, k_2), k_{m-1} \in \mathcal{C}_{n,m-1}(k_1, \dots, k_{m-2}) \right\}, \quad (2.11)$$

where  $\mathcal{C}_{n,1} = \left\{ 0 \leq k_1 \leq (N_1(X) - N_1^*(X)) \wedge (N_1(Y) - N_1^*(Y)) \right\}$  and for  $i = 2, \dots, m-1$ ,

$$\mathcal{C}_{n,i}(k_1, \dots, k_{i-1}) = \left\{ 0 \leq k_i \leq (N_i(X) - N_i^*(X)) \wedge (N_i(Y) - N_i^*(Y)) \right\}, \quad (2.12)$$

and where

$$R = \sum_{i=1}^{m-1} \sum_{j=N_i^*+1}^{N_i^*+k_i} N_m^{T_i^{j-1}, T_i^j},$$

with the various  $N$ 's and  $T$ 's given above by (2.1), (2.2), (2.3) and (2.8).

The representation (2.10) has the great advantage of (essentially) only involving the quantities  $N_i$ ,  $N_i^*$ ,  $i = 1, 2, \dots, m-1$  and  $T_i^j$ ,  $i = 1, 2, \dots, m-1$ ,  $j = 1, 2, \dots$ , and  $N_m$ .

### 3 Probability

Let us now bring our probabilistic framework into the picture by first studying the random variables  $N_m^{T_i^{j-1}, T_i^j}$ ,  $i = 1, 2, \dots, m-1$  and  $j = 1, 2, \dots$  and then the random variables  $N_i^*$ ,  $i = 1, 2, \dots, m-1$ , appearing in  $R$  in (2.9).

**Proposition 3.1** *Let  $(Z_n)_{n \geq 1}$  be a sequence of iid random variables with  $\mathbb{P}(Z_1 = \alpha_i) = p_i$ ,  $i = 1, \dots, m$ . For each  $i = 1, 2, \dots, m$ , let  $T_i^0 = 0$ , and let  $T_i^j$ ,  $j = 1, 2, \dots$  be the location of the  $j^{\text{th}}$   $\alpha_i$  in the infinite sequence  $(Z_n)_{n \geq 1}$ . Let  $i, r \in \{1, \dots, m\}$ , with  $r \neq i$ . Then, for any  $j = 1, 2, \dots$ , the conditional law of  $N_r^{T_i^{j-1}, T_i^j}$  given  $(T_i^{j-1}, T_i^j)$ , is binomial with parameters  $T_i^j - T_i^{j-1} - 1$  and  $p_r/(1-p_i)$ , which we denote by  $\mathcal{B}(T_i^j - T_i^{j-1} - 1, p_r/(1-p_i))$ . Moreover, the conditional law of  $(N_r^{T_i^{j-1}, T_i^j})_{r=1, \dots, m, r \neq i}$  given  $(T_i^{j-1}, T_i^j)$ , is multinomial with parameters  $T_i^j - T_i^{j-1} - 1$  and  $(p_r/(1-p_i))_{r=1, \dots, m, r \neq i}$ , which we denote by  $\mathcal{Mul}(T_i^j - T_i^{j-1} - 1, (p_r/(1-p_i))_{r=1, \dots, m, r \neq i})$ . Finally, for each  $i \neq r$ , the random variables  $(N_r^{T_i^{j-1}, T_i^j})_{j \geq 1}$ , are independent with mean  $p_r/p_i$  and variance  $(p_r/p_i)(1 + p_r/p_i)$ ; and, moreover, they are identically distributed in case the  $(Z_n)_{n \geq 1}$ , are uniformly distributed.*



**Proof.** Let us denote by  $\mathcal{L}(N_r^{T_i^{j-1}, T_i^j} | T_i^{j-1}, T_i^j)$  the conditional law of  $N_r^{T_i^{j-1}, T_i^j}$  given  $T_i^{j-1}, T_i^j$ . Recall, see (2.3), that  $T_i^{j-1}$  and  $T_i^j$  are the respective locations of the  $(j-1)^{\text{th}}$   $\alpha_i$  and the  $j^{\text{th}}$   $\alpha_i$  in the infinite sequence  $(Z_n)_{n \geq 1}$ . Thus between  $T_i^{j-1} + 1$  and  $T_i^j$ , there are  $T_i^j - T_i^{j-1} - 1$  free spots and each one is equally likely contain  $\alpha_r$ ,  $r \neq i$ , with probability  $p_r / (\sum_{\ell=1}^m p_\ell) = p_r / (1 - p_i)$ . Therefore,

$$\mathcal{L}\left(N_r^{T_i^{j-1}, T_i^j} | T_i^{j-1}, T_i^j\right) = \mathcal{B}\left(T_i^j - T_i^{j-1} - 1, \frac{p_r}{1 - p_i}\right). \quad (3.1)$$

Let us now compute the probability generating function of the random variables  $N_r^{T_i^{j-1}, T_i^j}$ ,  $i \neq r$ . First, via (3.1)

$$\begin{aligned} \mathbb{E}\left[x^{N_r^{T_i^{j-1}, T_i^j}}\right] &= \mathbb{E}\left[\mathbb{E}\left[x^{N_r^{T_i^{j-1}, T_i^j}} | T_i^{j-1}, T_i^j\right]\right] \\ &= \sum_{\ell=1}^{\infty} \left(1 - \frac{p_r}{1 - p_i} + \frac{p_r}{1 - p_i} x\right)^{\ell-1} p_i (1 - p_i)^{\ell-1} \\ &= \frac{p_i}{1 - (1 - p_i) \left(1 - \frac{p_r}{1 - p_i} + \frac{p_r}{1 - p_i} x\right)} \\ &= \frac{p_i}{p_i + p_r - p_r x}, \end{aligned} \quad (3.2)$$

since  $T_i^j$  is a negative binomial (Pascal) random variable with parameters  $j$  and  $p_i$  which we shall denote  $\mathcal{BN}(j, p_i)$  in the sequel and  $T_i^j - T_i^{j-1}$  is a geometric random variables with parameter  $p_i$ , which we shall denote  $\mathcal{G}(p_i)$ . Therefore,

$$\begin{aligned} \mathbb{E}\left[N_r^{T_i^{j-1}, T_i^j}\right] &= \frac{p_r}{p_i}, \\ \text{Var}\left(N_r^{T_i^{j-1}, T_i^j}\right) &= \frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right). \end{aligned} \quad (3.3)$$

In the uniform case, i.e.,  $p_i = 1/m$ ,  $i = 1, \dots, m$ , the  $N_r^{T_i^{j-1}, T_i^j}$ ,  $i = 1, \dots, m$ ,  $i \neq r$ ,  $j = 1, 2, \dots$  are clearly seen to be identically distributed, via (3.2). The multinomial part of the statement is proved in a very similar manner. The  $T_i^j - T_i^{j-1} - 1$  free spots are to contain the letters  $\alpha_r$ ,  $r \in \{1, \dots, m\}$ ,  $r \neq i$ , with respective probabilities  $p_r / (1 - p_i)$ . Therefore,

$$\mathcal{L}\left((N_r^{T_i^{j-1}, T_i^j})_{r=1, \dots, m, r \neq i} | T_i^{j-1}, T_i^j\right) = \mathcal{Mul}\left(T_i^j - T_i^{j-1} - 1, \left(\frac{p_r}{1 - p_i}\right)_{r=1, \dots, m, r \neq i}\right). \quad (3.4)$$

Via (3.4), the probability generating function of the random vector  $(N_r^{T_i^{j-1}, T_i^j})_{r=1, \dots, m, r \neq i}$  is then given by:

$$\mathbb{E}\left[\prod_{r=1, r \neq i}^m x_r^{N_r^{T_i^{j-1}, T_i^j}}\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{r=1, r \neq i}^m x_r^{N_r^{T_i^{j-1}, T_i^j}} | T_i^{j-1}, T_i^j\right]\right]$$

$$\begin{aligned}
&= \sum_{\ell=1}^{\infty} \left( \sum_{\substack{r=1 \\ r \neq i}}^m \frac{p_r}{1-p_i} x_r \right)^{\ell-1} p_i (1-p_i)^{\ell-1} \\
&= \frac{p_i}{1 - \sum_{r=1, r \neq i}^m p_r x_r}.
\end{aligned} \tag{3.5}$$

As a direct consequence of (3.5) and for  $r \neq i, s \neq i$ ,

$$\text{Cov} \left( N_r^{T_i^{j-1}, T_i^j}, N_s^{T_i^{j-1}, T_i^j} \right) = \frac{p_r p_s}{p_i^2}.$$

The proof of the proposition will be complete once, for each  $i \neq r$ , the random variables  $N_r^{T_i^{j-1}, T_i^j}$ ,  $j \geq 1$ , are shown to be independent. First, note that given  $T_i^{j-1}, T_i^j, T_i^{k-1}, T_i^k$ , the random variables  $N_r^{T_i^{j-1}, T_i^j} = \sum_{\ell=T_i^{j-1}+1}^{T_i^j} \mathbf{1}_{\{X_\ell=\alpha_r\}}$  and  $N_r^{T_i^{k-1}, T_i^k} = \sum_{\ell=T_i^{k-1}+1}^{T_i^k} \mathbf{1}_{\{X_\ell=\alpha_r\}}$  are independent since the intervals  $[T_i^{j-1}+1, T_i^j]$  and  $[T_i^{k-1}+1, T_i^k]$  are disjoint, and since the  $(X_\ell)_{\ell \geq 1}$  are also independent. Moreover, recall that conditional distributions are given by (3.1), and so, for instance,

$$\begin{aligned}
\mathcal{L} \left( N_r^{T_i^{j-1}, T_i^j} \mid T_i^{j-1}, T_i^j, T_i^{k-1}, T_i^k \right) &= \mathcal{L} \left( N_r^{T_i^{j-1}, T_i^j} \mid T_i^{j-1}, T_i^j \right) \\
&= \mathcal{B} \left( T_i^j - T_i^{j-1} - 1, \frac{p_r}{1-p_i} \right).
\end{aligned}$$

Therefore, for any measurable functions  $f, g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , and if  $\mathbb{E}_{\mathcal{B}(n,p)}$  denotes the expectation with respect to a binomial  $\mathcal{B}(n, p)$  distribution then

$$\begin{aligned}
&\mathbb{E} \left[ f(N_r^{T_i^{j-1}, T_i^j}) g(N_r^{T_i^{k-1}, T_i^k}) \right] \\
&= \mathbb{E} \left[ \mathbb{E} \left[ f(N_r^{T_i^{j-1}, T_i^j}) g(N_r^{T_i^{k-1}, T_i^k}) \mid T_i^{j-1}, T_i^j, T_i^{k-1}, T_i^k \right] \right] \\
&= \mathbb{E} \left[ \mathbb{E} \left[ f(N_r^{T_i^{j-1}, T_i^j}) \mid T_i^{j-1}, T_i^j, T_i^{k-1}, T_i^k \right] \mathbb{E} \left[ g(N_r^{T_i^{k-1}, T_i^k}) \mid T_i^{j-1}, T_i^j, T_i^{k-1}, T_i^k \right] \right] \tag{3.6}
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[ \mathbb{E}_{\mathcal{B}(T_i^j - T_i^{j-1} - 1, \frac{p_r}{1-p_i})} [f] \mathbb{E}_{\mathcal{B}(T_i^k - T_i^{k-1} - 1, \frac{p_r}{1-p_i})} [g] \right] \\
&= \mathbb{E} \left[ \mathbb{E}_{\mathcal{B}(T_i^j - T_i^{j-1} - 1, \frac{p_r}{1-p_i})} [f] \right] \mathbb{E} \left[ \mathbb{E}_{\mathcal{B}(T_i^k - T_i^{k-1} - 1, \frac{p_r}{1-p_i})} [g] \right] \tag{3.7} \\
&= \mathbb{E} \left[ f(N_r^{T_i^{j-1}, T_i^j}) \right] \mathbb{E} \left[ g(N_r^{T_i^{k-1}, T_i^k}) \right],
\end{aligned}$$

where the equality in (3.6) is due to the conditional independence property, while the one in (3.7) follows from that

$$\mathbb{E}_{\mathcal{B}(T_i^j - T_i^{j-1} - 1, \frac{p_r}{1-p_i})} [f] = F(T_i^j - T_i^{j-1}) \quad \text{and} \quad \mathbb{E}_{\mathcal{B}(T_i^k - T_i^{k-1} - 1, \frac{p_r}{1-p_i})} [g] = G(T_i^k - T_i^{k-1}),$$

for some functions  $F, G$ , and from the independence of  $T_i^j - T_i^{j-1}$  and  $T_i^k - T_i^{k-1}$ . The argument can then be easily adapted to justify the mutual independence of the random variables  $(N_r^{T_i^{j-1}, T_i^j})_{j \geq 1}$ .  $\square$

With the help of the previous proposition and in order to prepare our first fluctuation result, it is relevant to rewrite the representation (2.10) as

$$\text{LCI}_n = \max_{c_n} \min \left\{ \sum_{i=1}^{m-1} k_i + N_m(X) - G_{n,m}(X), \sum_{i=1}^{m-1} k_i + N_m(Y) - G_{n,m}(Y) \right\}, \quad (3.8)$$

where

$$G_{n,m} = \sum_{i=1}^{m-1} \sum_{j=N_i^*+1}^{N_i^*+k_i} \left( \left( \frac{N_m^{T_i^{j-1}, T_i^j} - \frac{p_m}{p_i}}{\sqrt{\frac{p_m}{p_i} \left(1 + \frac{p_m}{p_i}\right) n}} \right) \sqrt{\frac{p_m}{p_i} \left(1 + \frac{p_m}{p_i}\right) n} + \frac{p_m}{p_i} \right), \quad (3.9)$$

and where  $p_i(X) = \mathbb{P}(X_1 = \alpha_i)$  and  $p_i(Y) = \mathbb{P}(Y_1 = \alpha_i)$ ,  $1 \leq i \leq m$ . Recall once more that  $G_{n,m}$  actually depends on  $k_1, \dots, k_{m-1}$  but that, for the sake of readability, this dependency is omitted from our notations, see Remark 2.1.

Via (3.8) and (3.9),  $\text{LCI}_n$  is now represented as a max/min over random constraints of random sums of randomly stopped independent random variables, except for the presence of  $N_m(X)$  and  $N_m(Y)$ . Our next result also represents, up to a small error term, both  $N_m(X)$  and  $N_m(Y)$  via the same random variables.

**Proposition 3.2** *For each  $i = 1, 2, \dots, m$ , and  $r \neq i$ ,*

$$N_r = \frac{p_r}{p_i} N_i + \sum_{j=1}^{N_i} \frac{\left( N_r^{T_i^{j-1}, T_i^j} - \frac{p_r}{p_i} \right)}{\sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right) n}} \sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right) n} + S_{i,r}^{(n)}, \quad (3.10)$$

where  $\lim_{n \rightarrow +\infty} S_{i,r}^{(n)} / \sqrt{n} = 0$ , in probability. In particular, for each  $r = 1, 2, \dots, m$ ,

$$N_r = np_r + \sum_{\substack{i=1 \\ i \neq r}}^m \sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right) n} p_i \sum_{j=1}^{N_i} \frac{\left( N_r^{T_i^{j-1}, T_i^j} - \frac{p_r}{p_i} \right)}{\sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right) n}} + \sum_{\substack{i=1 \\ i \neq r}}^m p_i S_{i,r}^{(n)}. \quad (3.11)$$

**Proof.** Let us start the proof of (3.10) by identifying the random variable  $S_{i,r}^{(n)}$  and show that, when scaled by  $\sqrt{n}$ , they converge in probability to zero. Clearly, for  $i = 1, \dots, m$ ,  $i \neq r$ ,

$$0 \leq S_{i,r}^{(n)} := N_r - \sum_{j=1}^{N_i} N_r^{T_i^{j-1}, T_i^j}.$$

In other words,  $S_{i,r}^{(n)}$  is the number of  $\alpha_r$  in the interval  $[T_i^* + 1, n]$ , where  $T_i^*$  is the location of the last  $\alpha_i$  in  $[1, n]$ . Therefore,

$$0 \leq S_{i,r}^{(n)} \leq n - T_i^* = n - (T_i^{N_i} \wedge n). \quad (3.12)$$

But,  $\mathbb{P}(T_i^* = n - k) = p_i(1 - p_i)^k$ ,  $k = 0, 1, \dots, n - 1$  and  $\mathbb{P}(T_i^* = 0) = (1 - p_i)^n$ . Hence, for all  $\epsilon > 0$ , and  $n$  large enough,

$$\mathbb{P}\left(\frac{S_{i,m}^{(n)}}{\sqrt{n}} \geq \epsilon\right) \leq \mathbb{P}(n - T_i^* \geq \epsilon\sqrt{n}) \leq \sum_{l=\lceil \epsilon\sqrt{n} \rceil}^n p_i(1 - p_i)^l \leq (1 - p_i)^{\lceil \epsilon\sqrt{n} \rceil} \xrightarrow{n \rightarrow +\infty} 0. \quad (3.13)$$

Let us continue with the proof of (3.11). Summing over  $i = 1, \dots, m$ ,  $i \neq r$ , both sides of (3.10), we get

$$\sum_{\substack{i=1 \\ i \neq r}}^m \frac{p_i}{p_r} N_r = \sum_{\substack{i=1 \\ i \neq r}}^m N_i + \sum_{\substack{i=1 \\ i \neq r}}^m \sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right)} n \frac{p_i}{p_r} \left( \sum_{j=1}^{N_i} \frac{\left(N_r^{T_i^{j-1}, T_i^j} - \frac{p_r}{p_i}\right)}{\sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right)} n} \right) + \sum_{\substack{i=1 \\ i \neq r}}^m \frac{p_i}{p_r} S_{i,r}^{(n)}. \quad (3.14)$$

But,  $\sum_{i=1}^m N_i = n$ , and so (3.14) becomes

$$N_r = np_r + \sum_{\substack{i=1 \\ i \neq r}}^m \sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right)} np_i \left( \sum_{j=1}^{N_i} \frac{\left(N_r^{T_i^{j-1}, T_i^j} - \frac{p_r}{p_i}\right)}{\sqrt{\frac{p_r}{p_i} \left(1 + \frac{p_r}{p_i}\right)} n} \right) + \sum_{\substack{i=1 \\ i \neq r}}^m p_i S_{i,r}^{(n)},$$

which is precisely (3.11).  $\square$

**Remark 3.1** For all  $i \neq r$ ,  $\lim_{n \rightarrow +\infty} \mathbb{E}[(S_{i,r}^{(n)})^2/n] = 0$ . Indeed,

$$\begin{aligned} \mathbb{E}[(S_{i,r}^{(n)})^2/n] &= \int_0^{+\infty} \mathbb{P}((S_{i,m}^{(n)})^2 \geq xn) dx \leq \int_0^{+\infty} (1 - p_i)^{\lceil \sqrt{xn} \rceil} dx \\ &\leq \int_0^{+\infty} (1 - p_i)^{\sqrt{xn}-1} dx = \frac{2}{n(1 - p_i)(\ln(1 - p_i))^2}. \end{aligned}$$

Returning to the representation (3.8), the previous proposition allows us to rewrite  $\text{LCI}_n$  as:

$$\text{LCI}_n = \max_{\bigcap_{i=1}^{m-1} \mathcal{C}_{n,i}} \min \left( np_m(X) + \sum_{i=1}^{m-1} k_i - p_m(X) \sum_{i=1}^{m-1} \frac{k_i}{p_i(X)} + H_{m,n}(X) + \sum_{i=1}^{m-1} p_i(X) S_{i,m}^{(n)}(X), \right.$$

$$np_m(Y) + \sum_{i=1}^{m-1} k_i - p_m(Y) \sum_{i=1}^{m-1} \frac{k_i}{p_i(Y)} + H_{m,n}(Y) + \sum_{i=1}^{m-1} p_i(Y) S_{i,m}^{(n)}(Y) \Bigg), \quad (3.15)$$

where omitting the dependency in  $k_1, \dots, k_{m-1}$  (see Remark 2.1),

$$H_{m,n} = \sum_{i=1}^{m-1} \sqrt{\frac{p_m}{p_i} \left(1 + \frac{p_m}{p_i}\right)} np_i \sum_{j=1}^{N_i} \frac{\left(N_m^{T_i^{j-1}, T_i^j} - \frac{p_m}{p_i}\right)}{\sqrt{\frac{p_m}{p_i} \left(1 + \frac{p_m}{p_i}\right)} n} - \sum_{i=1}^{m-1} \sqrt{\frac{p_m}{p_i} \left(1 + \frac{p_m}{p_i}\right)} n \sum_{j=N_i^*+1}^{N_i^*+k_i} \frac{\left(N_m^{T_i^{j-1}, T_i^j} - \frac{p_m}{p_i}\right)}{\sqrt{\frac{p_m}{p_i} \left(1 + \frac{p_m}{p_i}\right)} n}. \quad (3.16)$$

We now study some of the properties of the random variables  $N_i^*$  which are present in both the random constraints and the random sums. The random variables  $N_i^*$  are defined recursively by (2.8) with  $N_1^* = 0$ . We fix  $\mathbf{k} = (k_1, \dots, k_{m-1})$  where  $k_i$  is the number of letters  $\alpha_i$  present in the common increasing subsequences. The random variables  $N_i^*$ ,  $i \geq 2$ , depend on  $\mathbf{k}$ , actually  $N_i^* = N_i^*(k_1, \dots, k_{i-1})$ . We write

$$N_i^* = \sum_{j=1}^{i-1} N_{i,j}^* \quad (3.17)$$

where  $N_{i,j}^* = N_{i,j}^*(k_j)$  is the number of letters  $\alpha_i$  present in the step  $j \leq i-1$  consisting in collecting the  $k_j$  letters  $\alpha_j$ ,  $j \leq i-1$ . (In the sequel, in order not to further burden the notations, we shall skip the symbols  $k_j$ ,  $j = 1, \dots, i-1$ , in  $N_i^*$  and  $N_{i,j}^*$ .) The following diagram encapsulates the drawing of the letters:

1	$T_1^{k_1}$	$T_2^{k_2+N_2^*}$	$T_3^{k_3+N_3^*}$	$\dots$	$T_{j-1}^{k_{j-1}+N_{j-1}^*}$	$T_j^{k_j+N_j^*}$	$\dots$	$T_{i-2}^{k_{i-2}+N_{i-2}^*}$	$T_{i-1}^{k_{i-1}+N_{i-1}^*}$
$k_1 \alpha_1$		$k_2 \alpha_2$							
$N_{2,1}^* \alpha_2$		$N_{3,2}^* \alpha_3$				$k_j \alpha_j$			
$N_{3,1}^* \alpha_3$			$k_3 \alpha_3$						
$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
$N_{i,1}^* \alpha_i$	$N_{i,2}^* \alpha_i$	$N_{i,3}^* \alpha_i$	$\dots$		$N_{i,j}^* \alpha_i$	$\dots$		$N_{i,i-1}^* \alpha_{i-1}$	

In Step  $j \leq i-1$ , there are  $T_j^{k_j+N_j^*} - T_{j-1}^{k_{j-1}+N_{j-1}^*}$  letters selected but  $k_j$  letters are  $\alpha_j$ ,  $N_{j+1,j}^*$  are  $\alpha_{j+1}$ ,  $\dots$ ,  $N_{i-1,j}^*$  are  $\alpha_{i-1}$ , (for  $j = i-1$ , there are also  $k_j$  letters  $\alpha_j$  but none of the others  $\alpha_{j+1}$ , etc).

Moreover, there are  $T_j^{k_j+N_j^*} - T_{j-1}^{k_{j-1}+N_{j-1}^*} - k_j - N_{j+1,j}^* - \dots - N_{i-1,j}^*$  possible spots ( $T_j^{k_j+N_j^*} - T_{j-1}^{k_{j-1}+N_{j-1}^*} - k_j$  in case  $j = i-1$ ) in which the probability of having a  $\alpha_i$  is  $p_{i,j} := p_i / (1 - p_j - \dots - p_{i-1})$ . Therefore, conditionally on

$$\mathcal{G}_{i,j}(\mathbf{k}) = \sigma \left( N_{j+1,j}^*, \dots, N_{i-1,j}^*, T_{j-1}^{k_{j-1}+N_{j-1}^*}, T_j^{k_j+N_j^*} \right),$$

(the  $\sigma$ -field generated by  $N_{j+1,j}^*, \dots, N_{i-1,j}^*, T_{j-1}^{k_{j-1}+N_{j-1}^*}, T_j^{k_j+N_j^*}$ ) it follows that

$$N_{i,j}^* \sim \mathcal{B}\left(T_j^{k_j+N_j^*} - T_{j-1}^{k_{j-1}+N_{j-1}^*} - k_j - N_{j+1,j}^* - \dots - N_{i-1,j}^*, p_{i,j}\right). \quad (3.18)$$

The two forthcoming propositions respectively characterize the laws of  $N_{i,j}^*$  and of  $N_i^*$ .

**Proposition 3.3** *For each  $i = 2, \dots, m$ , the probability generating function of  $N_{i,j}^*$ ,  $1 \leq j \leq i-1$ , is given by*

$$\mathbb{E}\left[x^{N_{i,j}^*}\right] = \left(\frac{p_j}{p_j + p_i - p_i x}\right)^{k_j}. \quad (3.19)$$

Therefore,  $N_{i,j}^*$  is distributed as  $\sum_{\ell=1}^{k_j} (G_\ell - 1)$ , where  $(G_\ell)_{1 \leq \ell \leq k_j}$  are independent with geometric law  $\mathcal{G}(p_j/(p_j + p_i))$  and so,

$$\mathbb{E}[N_{i,j}^*] = \frac{p_i}{p_j} k_j \quad \text{and} \quad \text{Var}(N_{i,j}^*) = \left(1 + \frac{p_i}{p_j}\right) \frac{p_i}{p_j} k_j. \quad (3.20)$$

**Proof.** Recall that, for  $N \sim \mathcal{B}(n, p)$ ,  $\mathbb{E}[x^N] = (1 - p + px)^n$  while, for  $N \sim \mathcal{G}(p)$ ,  $\mathbb{E}[x^N] = px/(1 - (1 - p)x)$ . Using (3.18), we then have for  $N = T_j^{k_j+N_j^*} - T_{j-1}^{k_{j-1}+N_{j-1}^*} - k_j - N_{j+1,j}^* - \dots - N_{i-1,j}^*$ ,

$$\begin{aligned} \mathbb{E}\left[x^{N_{i,j}^*}\right] &= \mathbb{E}\left[\mathbb{E}\left[x^{N_{i,j}^*} | N\right]\right] \\ &= \mathbb{E}\left[(1 - p_{i,j} + p_{i,j}x)^{T_j^{k_j+N_j^*} - T_{j-1}^{k_{j-1}+N_{j-1}^*} - k_j - N_{j+1,j}^* - \dots - N_{i-1,j}^*}\right] \\ &= \mathbb{E}\left[y^{U-V}\right], \end{aligned} \quad (3.21)$$

setting  $y = (1 - p_{i,j} + p_{i,j}x)$ , and

$$U := T_j^{k_j+N_j^*} - T_{j-1}^{k_{j-1}+N_{j-1}^*} - k_j \sim \mathcal{BN}(k_j, p_j) * \delta_{-k_j} \quad (3.22)$$

$$V := \sum_{r=j+1}^{i-1} N_{r,j}^* \sim \mathcal{B}\left(U, \sum_{r=j+1}^{i-1} \frac{p_r}{1 - p_j}\right), \quad (3.23)$$

where for  $j = i-1$ , we also set  $V = 0$ . The notation  $\mathcal{BN}(k, p)$  above stands for the negative binomial (Pascal) distribution with parameters  $k$  and  $p$ . The parameters of the binomial random variables  $V$  in (3.23) stem from that  $V$  counts the number of letters  $\alpha_r$ ,  $j+1 \leq r \leq i-1$ , between two letters  $\alpha_j$ , while exactly  $k_j$  such letters are obtained, so that each  $\alpha_r$  has probability  $p_r/(1 - p_j)$  to appear. Hence,

$$\begin{aligned} \mathbb{E}\left[y^{U-V}\right] &= \mathbb{E}\left[\mathbb{E}\left[y^{U-V} | U\right]\right] \\ &= \mathbb{E}\left[y^U \mathbb{E}\left[y^{-V} | U\right]\right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[ y^U \left( 1 - \sum_{r=j+1}^{i-1} \frac{p_r}{1-p_j} + \frac{\sum_{r=j+1}^{i-1} p_r}{(1-p_j)y} \right)^U \right] \\
&= \mathbb{E} \left[ \left( \left( 1 - \sum_{r=j+1}^{i-1} \frac{p_r}{1-p_j} \right) y + \sum_{r=j+1}^{i-1} \frac{p_r}{1-p_j} \right)^{G_1-1} \right]^{k_j},
\end{aligned}$$

since, from (3.22),  $U \sim \sum_{\ell=1}^{k_j} (G_\ell - 1)$ , where the  $G_\ell$ ,  $1 \leq \ell \leq k_j$ , are iid with distribution  $\mathcal{G}(p_j)$ . Finally,

$$\begin{aligned}
\mathbb{E} \left[ y^{U-V} \right] &= \left( \frac{p_j}{1 - (1-p_j) \left( \left( 1 - \sum_{r=j+1}^{i-1} \frac{p_r}{1-p_j} \right) y + \sum_{r=j+1}^{i-1} \frac{p_r}{1-p_j} \right)} \right)^{k_j} \\
&= \left( \frac{p_j}{p_j + p_i - p_i x} \right)^{k_j},
\end{aligned}$$

since  $p_{i,j} = p_i / (1 - \sum_{r=j}^{i-1} p_r)$ . The expressions for the expectation and for the variance in (3.20) follow from straightforward computations.  $\square$

Recall that by convention,  $N_1^* = 0$ , and for  $2 \leq i \leq m$ , the following proposition gives the law of  $N_i^*$ :

**Proposition 3.4** *For each  $i = 2, \dots, m$ , the random variables  $(N_{i,j}^*)_{1 \leq j \leq i-1}$  are independent. Hence, the probability generating function of  $N_i^*$  is given by*

$$\mathbb{E} \left[ x^{N_i^*} \right] = \prod_{j=1}^{i-1} \left( \frac{p_j}{p_j + p_i - p_i x} \right)^{k_j}, \quad (3.24)$$

and so,

$$\mathbb{E}[N_i^*] = \sum_{j=1}^{i-1} \frac{p_i}{p_j} k_j \quad \text{and} \quad \text{Var}(N_i^*) = \sum_{j=1}^{i-1} \left( 1 + \frac{p_i}{p_j} \right) \frac{p_i}{p_j} k_j. \quad (3.25)$$

**Proof.** In view of Proposition 3.3 and of (3.17), it is enough to prove the first part of the proposition, i.e., to prove that the random variables  $N_{i,j}^*$ ,  $1 \leq j \leq i-1$ , are independent. In order to simplify notations, we only show that  $N_{i,1}^*$  and  $N_{i,2}^*$  are independent, but the argument can easily be extended to prove the full independence property. Since the  $T_i^k$ 's are stopping times, by the strong Markov property, observe that  $\sigma(X_1, \dots, X_{T_1^{k_1}}) \perp\!\!\!\perp \sigma(X_{T_1^{k_1}+1}, \dots, X_{T_2^{k_2}+N_2^*})$  where, again  $\sigma(X_1, \dots, X_n)$  denotes the  $\sigma$ -field generated by the random variables  $X_1, \dots, X_n$ , while  $\perp\!\!\!\perp$  stands for independence conditionally on  $T_1^{k_1}$ . Moreover,  $T_1^{k_1}$  and  $\sigma(X_{T_1^{k_1}+1}, \dots, X_{T_2^{k_2}+N_2^*})$  are independent, and thus so

are  $\sigma(X_1, \dots, X_{T_1^{k_1}})$  and  $\sigma(X_{T_1^{k_1}+1}, \dots, X_{T_2^{k_2+N_2^*}})$ . The independence of  $N_{i,1}^*$  and  $N_{i,2}^*$  becomes clear, since  $N_{i,1}^*$  is  $\sigma(X_1, \dots, X_{T_1^{k_1}})$ -measurable while  $N_{i,2}^*$  is  $\sigma(X_{T_1^{k_1}+1}, \dots, X_{T_2^{k_2+N_2^*}})$ -measurable. The whole conclusion of the proposition then follows.  $\square$

## 4 The Uniform Case

In this section, we specialize our results to the case where the letters are uniformly drawn from the alphabet, i.e.,  $p_i(X) = p_i(Y) = 1/m$ , for all  $1 \leq i \leq m$ . Hence, the functional  $\text{LCI}_n$  in (3.15) rewrites as

$$\text{LCI}_n = \max_{c_n} \min \left( \frac{n}{m} + H_{m,n}(X) + \frac{1}{m} \sum_{i=1}^{m-1} S_{i,m}^{(n)}(X), \frac{n}{m} + H_{m,n}(Y) + \frac{1}{m} \sum_{i=1}^{m-1} S_{i,m}^{(n)}(Y) \right), \quad (4.1)$$

and therefore

$$\begin{aligned} \frac{\text{LCI}_n - n/m}{\sqrt{2n}} &= \max_{c_n} \min \left( \frac{H_{m,n}(X)}{\sqrt{2n}} + \frac{1}{m\sqrt{2n}} \sum_{i=1}^{m-1} S_{i,m}^{(n)}(X), \right. \\ &\quad \left. \frac{H_{m,n}(Y)}{\sqrt{2n}} + \frac{1}{m\sqrt{2n}} \sum_{i=1}^{m-1} S_{i,m}^{(n)}(Y) \right). \end{aligned} \quad (4.2)$$

The following simple inequality, a version of which is already present in [HLM], will be of multiple use (see Appendix A.1 for a proof):

**Lemma 4.1** *Let  $a_k, b_k, c_k, d_k$ ,  $1 \leq k \leq K$ , be reals. Then,*

$$\left| \max_{k=1,\dots,K} (a_k \wedge b_k) - \max_{k=1,\dots,K} ((a_k + c_k) \wedge (b_k + d_k)) \right| \leq \max_{k=1,\dots,K} (|c_k| \vee |d_k|). \quad (4.3)$$

The previous lemma entails

$$\begin{aligned} &\left| \max_{c_n} \min \left( \frac{H_{m,n}(X)}{\sqrt{2n}} + \frac{1}{m\sqrt{2n}} \sum_{i=1}^{m-1} S_{i,m}^{(n)}(X), \frac{H_{m,n}(Y)}{\sqrt{2n}} + \frac{1}{m\sqrt{2n}} \sum_{i=1}^{m-1} S_{i,m}^{(n)}(Y) \right) \right. \\ &\quad \left. - \max_{c_n} \min \left( \frac{H_{m,n}(X)}{\sqrt{2n}}, \frac{H_{m,n}(Y)}{\sqrt{2n}} \right) \right| \\ &\leq \frac{1}{m\sqrt{2n}} \left( \left| \sum_{i=1}^{m-1} S_{i,m}^{(n)}(X) \right| \vee \left| \sum_{i=1}^{m-1} S_{i,m}^{(n)}(Y) \right| \right). \end{aligned}$$

But, from Proposition 3.2, as  $n \rightarrow +\infty$ , both  $S_{i,m}^{(n)}(X)/\sqrt{n} \xrightarrow{\mathbb{P}} 0$  and  $S_{i,m}^{(n)}(Y)/\sqrt{n} \xrightarrow{\mathbb{P}} 0$ , for all  $1 \leq i \leq m-1$  (see (3.13)). Therefore, the fluctuations of  $\text{LCI}_n$  expressed in (4.2) are the same as that of

$$\max_{c_n} \min \left( \frac{H_{m,n}(X)}{\sqrt{2n}}, \frac{H_{m,n}(Y)}{\sqrt{2n}} \right).$$



For uniform draws, the functional  $H_{m,n}$  in (3.16) rewrites as

$$\begin{aligned} H_{m,n} &= \sum_{i=1}^{m-1} \sqrt{2n} \frac{1}{m} \sum_{j=1}^{N_i} \frac{N_m^{T_i^{j-1}, T_i^j} - 1}{\sqrt{2n}} - \sum_{i=1}^{m-1} \sqrt{2n} \sum_{j=N_i^*+1}^{N_i^*+k_i} \frac{N_m^{T_i^{j-1}, T_i^j} - 1}{\sqrt{2n}} \\ &= \sqrt{2n} \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i)} \left( \frac{N_i}{n} \right) - \sum_{i=1}^{m-1} \left( B_n^{(i)} \left( \frac{N_i^* + k_i}{n} \right) - B_n^{(i)} \left( \frac{N_i^*}{n} \right) \right) \right), \end{aligned}$$

where  $B_n^{(i)}$  is the Brownian approximation defined from the random variables  $N_m^{T_i^{j-1}, T_i^j}$ ,  $j \geq 1$ , which are iid, by Proposition 3.1, centered and scaled to have variance one, i.e.,  $B_n^{(i)}$  is the polygonal process on  $[0, 1]$  defined by linear interpolation between the values

$$B_n^{(i)} \left( \frac{k}{n} \right) = \sum_{j=1}^k \frac{Z_j^{(i)}}{\sqrt{n}}, \quad (4.4)$$

where

$$Z_j^{(i)} = \frac{N_m^{T_i^{j-1}, T_i^j} - 1}{\sqrt{2}}. \quad (4.5)$$

Next, we present some heuristic arguments which provide the limiting behavior of

$$\begin{aligned} \max_{c_n} \min & \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i), X} \left( \frac{N_i(X)}{n} \right) - \sum_{i=1}^{m-1} \left( B_n^{(i), X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i), X} \left( \frac{N_i^*(X)}{n} \right) \right) \right. \\ & \left. \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i), Y} \left( \frac{N_i(Y)}{n} \right) - \sum_{i=1}^{m-1} \left( B_n^{(i), Y} \left( \frac{N_i^*(Y) + k_i}{n} \right) - B_n^{(i), Y} \left( \frac{N_i^*(Y)}{n} \right) \right) \right) \quad (4.6) \end{aligned}$$

knowing that, by Donsker theorem,  $(B_n^{(1)}, \dots, B_n^{(m-1)}) \xrightarrow{(C_0([0,1]))^{m-1}} (B^{(1)}, \dots, B^{(m-1)})$ ,  $n \rightarrow +\infty$ , where  $(B^{(1)}, \dots, B^{(m-1)})$  is a drift-less,  $(m-1)$ -dimensional, correlated Brownian motion on  $[0, 1]$ , which is also zero at the origin. The correlation structure of this multivariate Brownian motion is given by that of the  $Z_j^{(i)}$ ,  $1 \leq i \leq m-1$ , which in turn is given by

Proposition 3.1. Above,  $\xrightarrow{(C_0([0,1]))^{m-1}}$  stands for the convergence in law in the product space of continuous function on  $[0, 1]$  vanishing at the origin. Since the multivariate Donsker theorem is crucial in our argument, we give a precise statement:

**Theorem 4.1 (Donsker)** *Let  $(Z_j)_{j \geq 1}$  be iid square integrable centered random vectors in  $\mathbb{R}^{m-1}$ ,  $m \geq 2$ , with covariance matrix  $\Sigma$ . Let  $(B_n)_{t \in [0,1]}$  be the polygonal process defined, for each  $n \geq 1$ , by*

$$B_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} Z_k + \frac{(nt - [nt])}{\sqrt{n}} Z_{[nt]+1}, \quad t \in [0, 1].$$

*Then  $B_n \xrightarrow{(C_0([0,1]))^{m-1}} B$  where  $B$  is a Brownian motion on  $[0, 1]^{m-1}$  with covariance matrix  $t\Sigma$  and where  $\xrightarrow{(C_0([0,1]))^{m-1}}$  stands for the convergence in law in the product space of continuous function on  $[0, 1]$  vanishing at the origin.)*

**Proof.** The multivariate Donsker theorem easily derives from the classical univariate one for which we refer, for instance to [Bil, Th. 8.2] and from the multivariate CLT as follows. Recall that the convergence  $B_n \xrightarrow{(C_0([0,1]))^{m-1}} B$  is equivalent to the convergence of finite-dimensional distributions of  $B_n$  to that of  $B$  and to the tightness of  $(B_n)_{n \geq 1}$  in  $(C_0([0,1]))^{m-1}$ . First, the multivariate CLT gives the convergence of the finite-dimensional distributions of  $(B_n^{(1)}(t), \dots, B_n^{(m-1)}(t))_{0 \leq t \leq 1}$  with a covariance structure given by that of the  $Z_1^{(i)}$ ,  $1 \leq i \leq m-1$ . Second, the tightness of  $(B_n^{(1)}(t), \dots, B_n^{(m-1)}(t))_{0 \leq t \leq 1}$  is obtained from that of its coordinates: since  $B_n^{(i)}$  is tight for each  $1 \leq i \leq m-1$  by the univariate Donsker theorem, for all  $\varepsilon > 0$ , there is a compact  $K_i$  of  $C_0([0,1])$ , the usual space of continuous functions on  $[0,1]$  vanishing at the origin, such that  $\sup_{n \geq 1} \mathbb{P}(B_n^{(i)} \notin K_i) < \varepsilon$  and we have

$$\sup_{n \geq 1} \mathbb{P}\left((B_n^{(1)}, \dots, B_n^{(m-1)}) \notin K_1 \times \dots \times K_{m-1}\right) \leq \sup_{n \geq 1} \sum_{i=1}^{m-1} \mathbb{P}(B_n^{(i)} \notin K_i) < (m-1)\varepsilon,$$

with  $K_1 \times \dots \times K_{m-1}$  compact of  $(C_0([0,1]))^{m-1}$  so that  $(B_n^{(1),X}, \dots, B_n^{(m-1),X})$  is tight in  $C_0([0,1])^{m-1}$ .  $\square$

## Heuristics

Roughly speaking, there are three limits to handle in (4.6):

1. The limit of the constraints in the maximum over  $\mathcal{C}_n$ ;
2. The limit of the linear terms:  $\sum_{i=1}^{m-1} B_n^{(i),X} \left( \frac{N_i(X)}{n} \right)$ ;
3. The limit of the increments:  $\sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i),X} \left( \frac{N_i^*(X)}{n} \right) \right)$ ;

and, similarly, for  $X$  replaced by  $Y$ . Below, the symbol  $\rightsquigarrow$  indicates an heuristic replacement or an heuristic limit, as  $n \rightarrow +\infty$ .

First Limit (to be treated last, in Section 4.3): Since  $\mathcal{C}_{n,i}(k_1, \dots, k_{i-1}) = \{\mathbf{k} = (k_1, \dots, k_{m-1}) : 0 \leq k_i \leq \min(N_i(X) - N_i^*(X), N_i(Y) - N_i^*(Y))\}$ , (and, again, with vacuous constraints in case either  $N_i^*(X) > n$  or  $N_i^*(Y) > n$ ) and from the concentration property of the  $N_i^*$ , we expect (with again  $k_0 = 0$ , and  $t_0 = 0$ , below):

$$\begin{aligned} \mathcal{C}_{n,i}(k_1, \dots, k_{i-1}) &\rightsquigarrow \left\{ \mathbf{k} = (k_1, \dots, k_{m-1}) : 0 \leq k_i \leq \left( \mathbb{E}[N_i(X)] - \sum_{j=1}^{i-1} k_j \right) \wedge \left( \mathbb{E}[N_i(Y)] - \sum_{j=1}^{i-1} k_j \right) \right\} \\ &= \left\{ \mathbf{k} = (k_1, \dots, k_{m-1}) : \frac{1}{n} \sum_{j=1}^{i-1} k_j \leq \frac{1}{n} \sum_{j=1}^i k_j \leq \frac{\mathbb{E}[N_i]}{n}, i = 1, \dots, m-1 \right\}. \end{aligned}$$

Hence, for  $\mathcal{C}_n$  defined in (2.11):

$$\mathcal{C}_n \rightsquigarrow \mathcal{V}\left(\frac{1}{m}, \dots, \frac{1}{m}\right),$$

where  $\mathcal{V}(p_1, \dots, p_{m-1}) = \{\mathbf{t} = (t_1, \dots, t_{m-1}) : t_i \geq 0, i = 1, \dots, m-1, t_1 \leq p_1, t_1 + t_2 \leq p_2, \dots, t_1 + \dots + t_{m-1} \leq p_{m-1}\}$ .

Second Limit (see Section 4.1): For each  $i = 1, \dots, m-1$ , the random variables  $N_i$  are concentrated around their respective mean  $\mathbb{E}[N_i] (= 1/m)$ , and so

$$\frac{N_i}{n} \rightsquigarrow \mathbb{E}[N_i] \quad \text{and} \quad \sum_{i=1}^{m-1} B_n^{(i)}\left(\frac{N_i}{n}\right) \rightsquigarrow \sum_{i=1}^{m-1} B^{(i)}(\mathbb{E}[N_i]) = \sum_{i=1}^{m-1} B^{(i)}\left(\frac{1}{m}\right),$$

where the limit  $B_n^{(i)} \xrightarrow{C_0([0,1])} B^{(i)}$  is taken simultaneously.

Third Limit (see Section 4.2): For each  $i = 1, \dots, m-1$ , the random variables  $N_i^*$  are also concentrated around their mean  $\mathbb{E}[N_i^*] = \sum_{j=1}^{i-1} k_j$ , and so  $N_i^* \rightsquigarrow \sum_{j=1}^{i-1} k_j$ . Therefore,

$$\begin{aligned} B_n^{(i),X}\left(\frac{N_i^*(X) + k_i}{n}\right) - B_n^{(i),X}\left(\frac{N_i^*(X)}{n}\right) &\rightsquigarrow B_n^{(i),X}\left(\sum_{j=1}^i \frac{k_j}{n}\right) - B_n^{(i),X}\left(\sum_{j=1}^{i-1} \frac{k_j}{n}\right) \\ &\rightsquigarrow B^{(i),X}\left(\sum_{j=1}^i t_j\right) - B^{(i),X}\left(\sum_{j=1}^{i-1} t_j\right), \end{aligned}$$

and similarly for  $X$  replaced by  $Y$ . Hence,

$$\begin{aligned} \frac{\text{LCI}_n - n/m}{\sqrt{2n}} &\rightsquigarrow \max_{\mathcal{V}(1/m, \dots, 1/m)} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B^{(i),X}\left(\frac{1}{m}\right) - \sum_{i=1}^{m-1} \left( B^{(i),X}\left(\sum_{j=1}^i t_j\right) - B^{(i),X}\left(\sum_{j=1}^{i-1} t_j\right) \right), \right. \\ &\quad \left. \frac{1}{m} \sum_{i=1}^{m-1} B^{(i),Y}\left(\frac{1}{m}\right) - \sum_{i=1}^{m-1} \left( B^{(i),Y}\left(\sum_{j=1}^i t_j\right) - B^{(i),Y}\left(\sum_{j=1}^{i-1} t_j\right) \right) \right) \\ &\stackrel{\mathcal{L}}{=} \frac{1}{\sqrt{m}} \max_{0=u_0 \leq u_1 \leq \dots \leq u_{m-1} \leq 1} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B^{(i),X}(1) - \sum_{i=1}^{m-1} (B^{(i),X}(u_i) - B^{(i),X}(u_{i-1})), \right. \\ &\quad \left. \frac{1}{m} \sum_{i=1}^{m-1} B^{(i),Y}(1) - \sum_{i=1}^{m-1} (B^{(i),Y}(u_i) - B^{(i),Y}(u_{i-1})) \right), \end{aligned}$$

by Brownian scaling and the reparametrization  $\sum_{j=1}^i t_j = u_i/m$ ,  $i = 1, \dots, m-1$ ,  $u_0 = t_0 = 0$ . In other words,

$$\frac{\text{LCI}_n - n/m}{\sqrt{2n/m}} \rightsquigarrow \max_{0=u_0 \leq u_1 \leq \dots \leq u_{m-1} \leq 1} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B^{(i),X}(1) - \sum_{i=1}^{m-1} (B^{(i),X}(u_i) - B^{(i),X}(u_{i-1})), \right.$$

$$\frac{1}{m} \sum_{i=1}^{m-1} B^{(i),Y}(1) - \sum_{i=1}^{m-1} (B^{(i),Y}(u_i) - B^{(i),Y}(u_{i-1})) \Bigg).$$

Finally, a linear transformation and Brownian properties allow to transform the parameter space into the Weyl chamber

$$\mathcal{W}_m(1) := \{\mathbf{t} = (t_0, t_1, \dots, t_{m-1}, t_m) : 0 = t_0 \leq t_1 \leq \dots \leq t_{m-1} \leq t_m = 1\},$$

and to replace the  $(m-1)$ -dimensional correlated Brownian motion  $B^X$  (resp.  $B^Y$ ), by an  $m$ -dimensional standard one  $B_1$  (resp.  $B_2$ ). Combining these facts, the expression on the right-hand side above, becomes equal, in law, to:

$$\begin{aligned} \max_{\mathbf{t} \in \mathcal{W}_m(1)} \min & \left( -\frac{1}{m} \sum_{i=1}^m B_1^{(i)}(1) + \sum_{i=1}^m (B_1^{(i)}(t_i) - B_1^{(i)}(u_{i-1})) \right), \\ & -\frac{1}{m} \sum_{i=1}^m B_2^{(i)}(1) + \sum_{i=1}^m (B_2^{(i)}(t_i) - B_2^{(i)}(u_{i-1})) \Bigg), \end{aligned}$$

which is the final form of our result, Theorem 1.1. In the sequel, we make precise the previous heuristic arguments.

All along, we use different sets constraints. For easy references, we gather here the references to these notations:  $\tilde{\mathcal{C}}_n$  is defined in (2.6),  $\tilde{\mathcal{C}}_{n,i}(k_1, \dots, k_{i-1})$  in (2.7),  $\mathcal{C}_n$  in (2.11),  $\mathcal{C}_{n,i}(k_1, \dots, k_{i-1})$  in (2.12),  $\mathcal{C}_{n,i}^*$  in (4.22),  $\mathcal{C}_n^*$  in (4.23),  $\mathcal{C}_{n,i}$  above (4.29),  $\mathcal{C}_{n,i}^\#$  in (4.29),  $\mathcal{C}_n^\pm$  in (4.59).

## 4.1 The Linear Terms

Set

$$R(X) = \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{N_i^* + k_i}{n} \right) - B_n^{(i),X} \left( \frac{N_i^*}{n} \right) \right),$$

where again the dependency of  $R(X)$  in  $(k_1, \dots, k_{m-1})$  is omitted (see Remark 2.1), so that with the help of (4.6), (4.2) rewrites as:

$$\begin{aligned} \frac{\text{LCI}_n - n/m}{\sqrt{2n}} &= \max_{\mathcal{C}_n} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X} \left( \frac{N_i(X)}{n} \right) - R(X), \right. \\ &\quad \left. \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y} \left( \frac{N_i(Y)}{n} \right) - R(Y) \right) + o_{\mathbb{P}}(1), \quad (4.7) \end{aligned}$$

where, throughout,  $o_{\mathbb{P}}(1)$  indicates a term, which might be different from an expression to another, converging to zero, in probability, as  $n$  converges to infinity.

Next, by Lemma 4.1,

$$\begin{aligned}
& \left| \max_{\mathcal{C}_n} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X} \left( \frac{N_i(X)}{n} \right) - R(X), \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y} \left( \frac{N_i(Y)}{n} \right) - R(Y) \right) \right. \\
& \quad \left. - \max_{\mathcal{C}_n} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X} \left( \frac{\mathbb{E}[N_i(X)]}{n} \right) - R(X), \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y} \left( \frac{\mathbb{E}[N_i(Y)]}{n} \right) - R(Y) \right) \right| \\
& \leq \max_{\mathcal{C}_n} \left| \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X} \left( \frac{N_i(X)}{n} \right) - R(X), \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y} \left( \frac{N_i(Y)}{n} \right) - R(Y) \right) \right. \\
& \quad \left. - \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X} \left( \frac{\mathbb{E}[N_i(X)]}{n} \right) - R(X), \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y} \left( \frac{\mathbb{E}[N_i(Y)]}{n} \right) - R(Y) \right) \right| \\
& \leq \max_{\mathcal{C}_n} \left( \max \left( \frac{1}{m} \left| \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{N_i(X)}{n} \right) - B_n^{(i),X} \left( \frac{\mathbb{E}[N_i(X)]}{n} \right) \right) \right|, \right. \right. \\
& \quad \left. \left. \frac{1}{m} \left| \sum_{i=1}^{m-1} \left( B_n^{(i),Y} \left( \frac{N_i(Y)}{n} \right) - B_n^{(i),Y} \left( \frac{\mathbb{E}[N_i(Y)]}{n} \right) \right) \right| \right) \right). \quad (4.8)
\end{aligned}$$

We now wish to show that the right-hand side of (4.8) converges to zero, in probability. First note that for each  $2 \leq i \leq m-1$ ,  $\mathcal{C}_{n,i}(k_1, \dots, k_{i-1}) \subset \{\mathbf{k} = (k_1, \dots, k_{m-1}) : 0 \leq k_i \leq \min(N_i(X), N_i(Y))\} \subset \{\mathbf{k} = (k_1, \dots, k_{m-1}) : 0 \leq k_i \leq n\}$ , and the same holds true for  $\mathcal{C}_{n,1}$ , see (2.11). But,  $B_n^{(i)}(N_i/n) - B_n^{(i)}(\mathbb{E}[N_i]/n)$ , where we have dropped  $X$  and  $Y$ , does not depend on  $\mathbf{k}$ . Therefore, the maximum can be skipped and the problem reduces to showing that, for all  $1 \leq i \leq m-1$ :

$$\left| B_n^{(i)} \left( \frac{N_i}{n} \right) - B_n^{(i)} \left( \frac{\mathbb{E}[N_i]}{n} \right) \right| \xrightarrow{\mathbb{P}} 0, \quad (4.9)$$

as  $n \rightarrow +\infty$ . This follows from the forthcoming lemma applied, for each  $i = 1, \dots, m-1$ , to the random variables  $Z_j^{(i)} = (N_m^{T_i^{j-1}, T_i^j} - 1)/\sqrt{2}$ , present in both (4.4) and (4.5) and which, by Proposition 3.1, are iid with mean zero and variance one. Note that the lemma below (see Appendix A.1 for a proof) can indeed be brought into play since Hoeffding's inequality, applied to the random variables  $N_i$ , ensures that for  $x_n = \sqrt{n} \ln n$ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|N_i - \mathbb{E}[N_i]| \geq x_n) \leq \lim_{n \rightarrow +\infty} 2e^{-2x_n^2/n} = 0. \quad (4.10)$$

**Lemma 4.2** *Let  $(Z_j)_{j \geq 1}$  be iid centered random variables with unit variance, and for each  $n \in \mathbb{N}$ , let  $N^{(n)}$  be an  $\mathbb{N}$ -valued random variable such that  $\lim_{n \rightarrow +\infty} \mathbb{P}(|N^{(n)} - \mathbb{E}[N^{(n)}]| \geq x_n) = 0$ , where  $x_n \geq 0$  is such that  $\lim_{n \rightarrow +\infty} x_n/n = 0$ . Then,*

$$\sum_{j \in [N^{(n)}, \mathbb{E}[N^{(n)}]]} \frac{Z_j}{\sqrt{n}} \xrightarrow{\mathbb{P}} 0,$$

where  $[N^{(n)}, \mathbb{E}[N^{(n)}]]$  is short for  $[\min(N^{(n)}, \mathbb{E}[N^{(n)}]), \max(N^{(n)}, \mathbb{E}[N^{(n)}])]$ .

At this stage, (4.9) is proved and therefore,

$$\begin{aligned} \frac{\text{LCI}_n - n/m}{\sqrt{2n}} &= \max_{c_n} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X} \left( \frac{\mathbb{E}[N_i(X)]}{n} \right) - R(X), \right. \\ &\quad \left. \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y} \left( \frac{\mathbb{E}[N_i(Y)]}{n} \right) - R(Y) \right) + o_{\mathbb{P}}(1), \end{aligned} \quad (4.11)$$

finishing the first part of the proof of Theorem 1.1. Indeed,  $(N_1, \dots, N_m)$  is multinomial with parameters  $n$  and  $(p_1, \dots, p_m)$ . So, for uniform draws,  $\mathbb{E}[N_i(X)] = \mathbb{E}[N_i(Y)] = np_i = n/m$ . Then, by the multivariate Donsker theorem, see Th. 4.1, and scaling,

$$\sum_{i=1}^{m-1} \frac{1}{m} B_n^{(i),X} \left( \frac{\mathbb{E}[N_i(X)]}{n} \right) \Rightarrow \sum_{i=1}^{m-1} \frac{1}{m\sqrt{m}} B^{(i),X}(1), \quad n \rightarrow +\infty, \quad (4.12)$$

where  $(B^{(1),X}(t), \dots, B^{(m-1),X}(t))_{0 \leq t \leq 1}$  is a  $(m-1)$ -dimensional Brownian motion and similarly for  $Y$ . As shown next, the covariance matrix of this Brownian motion at time  $t$  is  $t\Sigma = t(\sigma_{k,l})_{1 \leq k, l \leq m-1}$ , where

$$\Sigma = \begin{pmatrix} 1 & 1/2 & \dots & 1/2 \\ 1/2 & 1 & 1/2 & \vdots \\ \vdots & & \ddots & 1/2 \\ 1/2 & \dots & 1/2 & 1 \end{pmatrix}. \quad (4.13)$$

Indeed,  $\Sigma$  in (4.13) is obtained as follows: First, since

$$(B_n^{(1),X}, \dots, B_n^{(m-1),X}) \xrightarrow{(C_0([0,1]))^{m-1}} (B^{(1),X}, \dots, B^{(m-1),X}),$$

while uniform integrability (see Lemma 4.3, below) entails

$$\lim_{n \rightarrow +\infty} \text{Cov}(B_n^{(k),X}(1), B_n^{(l),X}(1)) = \text{Cov}(B^{(k),X}(1), B^{(l),X}(1)) = \sigma_{k,l}.$$

Next, in the uniform case, Proposition 3.2 writes, for  $i = 1, \dots, m-1$ , as

$$N_m = N_i + \sqrt{2n} B_n^{(i),X} \left( \frac{N_i}{n} \right) + o_{\mathbb{P}}(\sqrt{n}),$$

so that using also Remark 3.1

$$\text{Cov} \left( B_n^{(k),X} \left( \frac{N_k}{n} \right), B_n^{(l),X} \left( \frac{N_l}{n} \right) \right) = \frac{1}{2n} \text{Cov}(N_m - N_k, N_m - N_l) + o(1). \quad (4.14)$$

But  $(N_1, \dots, N_m) \sim \text{Mult}(n, (\frac{1}{m}, \dots, \frac{1}{m}))$ ,  $N_i/n \rightarrow 1/m$ ,  $\text{Var}(N_i) = n(m-1)/m^2$  and when,  $i \neq j$ ,  $\text{Cov}(N_i, N_j) = -n/m^2$ . Therefore,

$$\frac{1}{2n} \text{Cov}(N_m - N_k, N_m - N_l) = \frac{1}{2n} \left( \frac{n(m-1)}{m^2} + \frac{n}{m^2} - \frac{n}{m^2} - \frac{n}{m^2} \right) = \frac{1}{2m}. \quad (4.15)$$

Since by Lemma 4.3,  $\lim_{n \rightarrow +\infty} \mathbb{E}[(B_n^{(k)}(N_k/n) - B_n^{(k)}(1/m))^2] = 0$ , it follows

$$\begin{aligned} \lim_{n \rightarrow +\infty} \text{Cov} \left( B_n^{(k),X} \left( \frac{N_k}{n} \right), B_n^{(l),X} \left( \frac{N_l}{n} \right) \right) &= \lim_{n \rightarrow +\infty} \text{Cov} \left( B_n^{(k),X} \left( \frac{1}{m} \right), B_n^{(l),X} \left( \frac{1}{m} \right) \right) \\ &= \text{Cov} \left( B^{(k),X} \left( \frac{1}{m} \right), B^{(l),X} \left( \frac{1}{m} \right) \right) \\ &= \frac{1}{m} \text{Cov} \left( B^{(k),X}(1), B^{(l),X}(1) \right). \end{aligned} \quad (4.16)$$

Finally, (4.14), (4.15), (4.16) ensure the expression (4.13) for the covariance. To finish, let us state a lemma, just used above and, whose proof is presented in Appendix A.1.

**Lemma 4.3** *The sequences  $(B_n^{(k)}(N_k/n)^2)_{n \geq 1}$  and  $(B_n^{(k)}(1/m)^2)_{n \geq 1}$ ,  $k = 1, \dots, m-1$ , are uniformly integrable and*

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[ \left( B_n^{(k)} \left( \frac{N_k}{n} \right) - B_n^{(k)} \left( \frac{1}{m} \right) \right)^2 \right] = 0. \quad (4.17)$$

## 4.2 The Increments

In this section, we compare the maximum of two different quantities over the same set of constraints in order to simplify the quantities to be maximized (before simplifying the constraints  $\mathcal{C}_n$  themselves, in the next section). The quantities to compare are:

$$\begin{aligned} \max_{\mathbf{k} \in \mathcal{C}_n} \left\{ \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X}(p_i(X)) - \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i),X} \left( \frac{N_i^*(X)}{n} \right) \right) \right) \wedge \right. \\ \left. \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y}(p_i(Y)) - \sum_{i=1}^{m-1} \left( B_n^{(i),Y} \left( \frac{N_i^*(Y) + k_i}{n} \right) - B_n^{(i),Y} \left( \frac{N_i^*(Y)}{n} \right) \right) \right) \right\} \end{aligned} \quad (4.18)$$

and

$$\begin{aligned} \max_{\mathbf{k} \in \mathcal{C}_n} \left\{ \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),X}(p_i(X)) - \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{\sum_{j=1}^i k_j}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right) \right) \wedge \right. \\ \left. \left( \frac{1}{m} \sum_{i=1}^{m-1} B_n^{(i),Y}(p_i(Y)) - \sum_{i=1}^{m-1} \left( B_n^{(i),Y} \left( \frac{\sum_{j=1}^i k_j}{n} \right) - B_n^{(i),Y} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right) \right) \right\} \end{aligned} \quad (4.19)$$

Using (4.3) in Lemma 4.1, their absolute difference is upper-bounded by

$$\begin{aligned} \max_{\mathbf{k} \in \mathcal{C}_n} \left\{ \left| \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i),X} \left( \frac{N_i^*(X)}{n} \right) \right) \right. \right. \\ \left. \left. - \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{\sum_{j=1}^i k_j}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right) \right| \right\} \end{aligned}$$

$$\begin{aligned}
& \bigvee \left| \sum_{i=1}^{m-1} \left( B_n^{(i),Y} \left( \frac{N_i^*(Y) + k_i}{n} \right) - B_n^{(i),Y} \left( \frac{N_i^*(Y)}{n} \right) \right. \right. \\
& \quad \left. \left. - \sum_{j=1}^{m-1} \left( B_n^{(i),Y} \left( \frac{\sum_{j=1}^i k_j}{n} \right) - B_n^{(i),Y} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right) \right| \right\} \\
& \leq \max_{\mathbf{k} \in \mathcal{C}_n} \left\{ \left| \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^i k_j}{n} \right) \right) \right| \right. \\
& \quad \bigvee \left| \sum_{i=1}^{m-1} \left( B_n^{(i),Y} \left( \frac{N_i^*(Y) + k_i}{n} \right) - B_n^{(i),Y} \left( \frac{\sum_{j=1}^i k_j}{n} \right) \right) \right| \Big\} \\
& + \max_{\mathbf{k} \in \mathcal{C}_n} \left\{ \left| \sum_{i=1}^{m-1} \left( B_n^{(i),X} \left( \frac{N_i^*(X)}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right) \right| \right. \\
& \quad \bigvee \left| \sum_{i=1}^{m-1} \left( B_n^{(i),Y} \left( \frac{N_i^*(Y)}{n} \right) - B_n^{(i),Y} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right) \right| \Big\}.
\end{aligned}$$

Recall that  $N_1^*(X) = N_1^*(Y) = 0$ . Hence, for  $i = 1$ ,

$$B_n^{(i),X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^i k_j}{n} \right) = B_n^{(i),X} \left( \frac{N_i^*(X)}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) = 0,$$

with the same property for functionals relative to  $Y$ . Therefore, we are left with investigating terms of the form

$$\begin{aligned}
& \max_{\mathbf{k} \in \mathcal{C}_n} \left\{ \left| B_n^{(i),X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^i k_j}{n} \right) \right| \right. \\
& \quad \left. \bigvee \left| B_n^{(i),Y} \left( \frac{N_i^*(Y) + k_i}{n} \right) - B_n^{(i),Y} \left( \frac{\sum_{j=1}^i k_j}{n} \right) \right| \right\}, \quad (4.20)
\end{aligned}$$

and

$$\max_{\mathbf{k} \in \mathcal{C}_n} \left\{ \left| B_n^{(i),X} \left( \frac{N_i^*(X)}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right| \bigvee \left| B_n^{(i),Y} \left( \frac{N_i^*(Y)}{n} \right) - B_n^{(i),Y} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right| \right\}, \quad (4.21)$$

for  $2 \leq i \leq m-1$ . Above, all the quantities considered only depend on a single sequence, say  $X$  or  $Y$ , except for the constraints in  $\mathcal{C}_n$  which depend on both  $X$  and  $Y$ . However,

$$\mathcal{C}_{n,i}(k_1, \dots, k_{i-1}) \subset \mathcal{C}_{n,i}^*(X) := \{\mathbf{k} = (k_1, \dots, k_{m-1}) : 0 \leq k_i \leq N_i(X) - N_i^*(X)\} \quad (4.22)$$

(resp.  $\mathcal{C}_{n,i}(k_1, \dots, k_{i-1}) \subset \mathcal{C}_{n,i}^*(Y)$ ), and the same for  $\mathcal{C}_{n,1}$  and so upper-bounding, in (4.20) and (4.21), the inner maxima by sums and the maxima over  $\mathcal{C}$  by maxima over

$$\mathcal{C}_n^*(X) := \bigcap_{i=1}^{m-1} \mathcal{C}_{n,i}^*(X), \quad (4.23)$$



(resp.  $\mathcal{C}_n^*(Y)$ ), we are left with investigating, for  $2 \leq i \leq m-1$ , the convergence in probability of terms of the form

$$\max_{\mathbf{k} \in \mathcal{C}_n^*(X)} \left\{ \left| B_n^{(i),X} \left( \frac{N_i^*(X) + k_i}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^i k_j}{n} \right) \right| \right\}, \quad (4.24)$$

and

$$\max_{\mathbf{k} \in \mathcal{C}_n^*(X)} \left\{ \left| B_n^{(i),X} \left( \frac{N_i^*(X)}{n} \right) - B_n^{(i),X} \left( \frac{\sum_{j=1}^{i-1} k_j}{n} \right) \right| \right\}, \quad (4.25)$$

and, similarly with  $X$  replaced by  $Y$ . Omitting the reference to either  $X$  or  $Y$ , the terms to control are, from (4.4) and for each,  $2 \leq i \leq m-1$ , of the form:

$$\max_{\mathbf{k} \in \mathcal{C}_n^*} \left| \sum_{j=k_1+\dots+k_{i-1}+1}^{N_i^*+k_i} \frac{Z_j^{(i)}}{\sqrt{n}} \right|, \quad (4.26)$$

and

$$\max_{\mathbf{k} \in \mathcal{C}_n^*} \left| \sum_{j=k_1+\dots+k_{i-1}+1}^{N_i^*} \frac{Z_j^{(i)}}{\sqrt{n}} \right|, \quad (4.27)$$

where the  $Z_j^{(i)}$ ,  $j \geq 1$ , are defined in (4.5) and where

$$\mathcal{C}_n^* = \bigcap_{i=1}^{m-1} \mathcal{C}_{n,i}^*, \quad \text{with } \mathcal{C}_{n,i}^* = \{ \mathbf{k} = (k_1, \dots, k_{m-1}) : 0 \leq k_i \leq N_i - N_i^* \}.$$

In (4.26), (4.27) and henceforth, we write  $\sum_{j=n_1}^{n_2}$  regardless of the order of  $n_1$  and  $n_2$ , i.e., by convention this sum is  $\sum_{j=n_2}^{n_1}$  when  $n_2 < n_1$ .

Since (4.27) is similar, but easier to tackle than (4.26), we only deal with (4.26). Again, as in Section 4.1, let  $D_n^i = \{ |N_i - \mathbb{E}[N_i]| \leq \sqrt{n} \ln n \}$  for  $i = 1, 2, \dots, m-1$ , and, thus, for  $\varepsilon > 0$ ,

$$\begin{aligned} & \mathbb{P} \left( \max_{\mathbf{k} \in \mathcal{C}_n^*} \left| \sum_{j=k_1+\dots+k_{i-1}+1}^{N_i^*+k_i} \frac{Z_j^{(i)}}{\sqrt{n}} \right| \geq \varepsilon \right) \\ & \leq \mathbb{P} \left( \left\{ \max_{\mathbf{k} \in \mathcal{C}_n^*} \left| \sum_{j=k_1+\dots+k_{i-1}+1}^{N_i^*+k_i} Z_j^{(i)} \right| \geq \varepsilon \sqrt{n} \right\} \cap \bigcap_{i=1}^{m-1} D_n^i \right) + \sum_{i=1}^{m-1} \mathbb{P}((D_n^i)^c). \end{aligned} \quad (4.28)$$

Let  $\mathcal{C}_n^{i-1} = \bigcap_{j=1}^{i-1} \{ k_j \leq \mathbb{E}[N_j] + \sqrt{n} \ln n \}$  and let  $\mathcal{C}_{n,i}^\#$  be the set of indices  $k_1, \dots, k_i$

$$\mathcal{C}_{n,i}^\# = \mathcal{C}_n^{i-1} \cap \left\{ \mathbb{E}[N_i^*] \leq \ell_i = k_1 + \dots + k_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n - (N_i^* - \mathbb{E}[N_i^*]) \right\} \quad (4.29)$$

where we set  $\ell_i := k_1 + \dots + k_i$ . Since under  $\bigcap_{i=1}^{m-1} D_n^i$ ,  $\mathcal{C}_{n,i}^* \subset \{k_j \leq \mathbb{E}[N_j] + \sqrt{n} \ln n\}$  and since Proposition 3.4, specialized to the uniform case, gives  $\mathbb{E}[N_i^*] = \sum_{j=1}^{i-1} k_j$ , it follows that  $\mathcal{C}_n^* \subset \mathcal{C}_{n,i}^\#$  and (4.28) is thus further upper-bounded by

$$\mathbb{P} \left( \max_{\mathbf{k} \in \mathcal{C}_{n,i}^\#} \left| \sum_{j=\ell_i+1}^{\ell_i+N_i^*-(k_1+\dots+k_{i-1})} Z_j^{(i)} \right| \geq \varepsilon \sqrt{n} \right) + \sum_{i=1}^{m-1} \mathbb{P}((D_n^i)^c). \quad (4.30)$$

Now, in view of (4.10), it is enough to show the convergence to zero of the first term on the right-hand side of (4.30). To do so, set  $E_n^1 = \Omega$  and, for  $2 \leq i \leq m-1$ ,

$$E_n^i(k_1, \dots, k_{i-1}) = \left\{ |N_i^*(k_1, \dots, k_{i-1}) - \mathbb{E}[N_i^*(k_1, \dots, k_{i-1})]| \leq x_n \right\},$$

with

$$x_n = \sqrt{n} \ln n, \quad (4.31)$$

and let

$$E_n^i = \bigcap_{(k_1, \dots, k_{i-1}) \in \mathcal{C}_n^{i-1}} E_n^i(k_1, \dots, k_{i-1}). \quad (4.32)$$

Our next goal is to show that asymptotically,  $E_n^i$  has full probability.

**Proposition 4.1** *Let  $2 \leq i \leq m-1$ , then  $\lim_{n \rightarrow +\infty} \mathbb{P}((E_n^i)^c) = 0$ .*

In order to prove Proposition 4.1, we first need the following technical result, proved in Appendix A.1:

**Lemma 4.4** *For  $x \in [-n, +\infty)$ , let*

$$K_n(x) = \frac{(x+2n)^{x+2n}}{(2x+2n)^{x+n}(2n)^n}.$$

*Then, for some constants  $c, C \in (0, +\infty)$ ,*

$$K_n(x) \leq C \exp \left( -cn \min \left( \frac{|x|}{n}, \frac{x^2}{n^2} \right) \right). \quad (4.33)$$

We proceed now to the proof of Proposition 4.1:

**Proof.** (Prop. 4.1) Clearly,

$$\begin{aligned} \mathbb{P}((E_n^i)^c) &\leq \sum_{(k_1, \dots, k_{i-1}) \in \mathcal{C}_n^{i-1}} \mathbb{P}((E_n^i(k_1, \dots, k_{i-1}))^c) \\ &\leq n^{i-1} \max_{(k_1, \dots, k_{i-1}) \in \mathcal{C}_n^{i-1}} \mathbb{P}((E_n^i(k_1, \dots, k_{i-1}))^c). \end{aligned}$$

Therefore, to prove the lemma, it is enough to show that:

$$\lim_{n \rightarrow +\infty} n^{i-1} \max_{(k_1, \dots, k_{i-1}) \in \mathcal{C}_n^{i-1}} \mathbb{P}((E_n^i(k_1, \dots, k_{i-1}))^c) = 0. \quad (4.34)$$

Now, for each  $2 \leq i \leq m-1$ , Propositions 3.3 and 3.4 assert that,

$$N_i^* = N_i^*(k_1, \dots, k_{i-1}) = \sum_{j=1}^{i-1} N_{i,j}^*,$$

where the  $(N_{i,j}^*)_{1 \leq j \leq i-1}$  are independent and with probability generating function

$$\mathbb{E}[x^{N_{i,j}^*}] = \left(\frac{1}{2-x}\right)^{k_j}.$$

Next,

$$\begin{aligned} \mathbb{P}(E_n^i(k_1, \dots, k_{i-1}))^c &= \mathbb{P}(|N_i^* - \mathbb{E}[N_i^*]| > x_n) \\ &= \mathbb{P}\left(\sum_{j=1}^{i-1} (N_{i,j}^* - k_j) > x_n\right) + \mathbb{P}\left(\sum_{j=1}^{i-1} (k_j - N_{i,j}^*) > x_n\right). \end{aligned} \quad (4.35)$$

The first term in (4.35) is bounded by  $\Theta_{k_1+\dots+k_{i-1}}^r(x_n)$ , where

$$\Theta_k^r(x) := \min_{t>0} \left( \exp \left( - (t(x+k) + k \ln(2-e^t)) \right) \right) \quad (4.36)$$

$$= \frac{(x+2k)^{x+2k}}{(2x+2k)^{x+k}(2k)^k}, \quad (4.37)$$

since the minimization in (4.36) occurs at  $t = \ln((2x+2k)/(x+2k))$ .

The second term in (4.35) is bounded by  $\Theta_{k_1+\dots+k_{i-1}}^l(x_n)$ , where

$$\Theta_k^l(x) := \min_{t>0} \left( \exp \left( - (t(x-k) + k \ln(2-e^{-t})) \right) \right) \quad (4.38)$$

$$= \frac{(2k-x)^{2k-x}}{(2k-2x)^{k-x}(2k)^k}, \quad (4.39)$$

observing that, for  $x \leq k$ , the minimization in (4.38) occurs at  $t = \ln((2k-x)/(2k-2x))$ .

From the previous bounds and (4.35), it is clear that (4.34) will follow from

$$\lim_{n \rightarrow +\infty} n^{i-1} \max_{(k_1, \dots, k_{i-1}) \in \mathcal{C}_n^{i-1}} \Theta_{k_1+\dots+k_{i-1}}^\bullet(x_n) = 0, \quad (4.40)$$

for  $\bullet \in \{l, r\}$ . To obtain such as limit, we make use of Lemma 4.4, with  $x = x_n = \sqrt{n} \ln(n)$ , noting also that  $\mathcal{C}_n^{i-1} \subset \{(k_1, \dots, k_{i-1}) : k_1 + \dots + k_{i-1} \leq \sum_{j=1}^{i-1} \mathbb{E}[N_j] + (i-1)\sqrt{n} \ln n\} \subset \{(k_1, \dots, k_{i-1}) : k_1 + \dots + k_{i-1} \leq (i-1)(\max_{j=1, \dots, i-1} p_j n + \sqrt{n} \ln n)\}$ .

First, for  $\bullet = r$ , when  $k_1 + \dots + k_{i-1} \leq x_n$ , (4.33) writes as

$$\Theta_{k_1+\dots+k_{i-1}}^r(x_n) \leq C \exp \left( - c(k_1 + \dots + k_{i-1}) \min \left( \frac{x_n}{k_1 + \dots + k_{i-1}}, \left( \frac{x_n}{k_1 + \dots + k_{i-1}} \right)^2 \right) \right)$$

$$= C \exp(-cx_n),$$

so that

$$n^{i-1} \max_{k_1+\dots+k_{i-1} \leq x_n} \Theta_{k_1+\dots+k_{i-1}}^r(x_n) \leq C n^{i-1} e^{-c\sqrt{n} \ln n} \rightarrow 0, \quad n \rightarrow +\infty,$$

where above, and below,  $C$  is a finite positive constant whose value might change from a line to another. For  $x_n \leq k_1 + \dots + k_{i-1} \leq (i-1)(n \max_{j=1,\dots,i-1} p_j + \sqrt{n} \ln n) = (i-1)(n/m + \sqrt{n} \ln n)$ , (4.33) writes as

$$\begin{aligned} \Theta_{k_1+\dots+k_{i-1}}^r(x_n) &\leq C \exp\left(-c(k_1 + \dots + k_{i-1}) \min\left(\frac{x_n}{k_1 + \dots + k_{i-1}}, \left(\frac{x_n}{k_1 + \dots + k_{i-1}}\right)^2\right)\right) \\ &= C \exp\left(-c \frac{x_n^2}{k_1 + \dots + k_{i-1}}\right) \\ &\leq C \exp\left(-c \frac{x_n^2}{(i-1)(n/m + \sqrt{n} \ln n)}\right), \end{aligned}$$

so that

$$\begin{aligned} n^{i-1} \max_{x_n \leq k_1+\dots+k_{i-1} \leq (i-1)(n/m + \sqrt{n} \ln n)} \Theta_{k_1+\dots+k_{i-1}}^r(x_n) \\ \leq n^{i-1} \exp\left(-c \frac{n(\ln n)^2}{(i-1)(n/m + \sqrt{n} \ln n)}\right) \rightarrow 0, \quad n \rightarrow +\infty, \end{aligned}$$

guaranteeing (4.40) with  $\bullet = r$ .

Next, let  $\bullet = l$  and consider the following three cases:  $k_1 + \dots + k_{i-1} \leq x_n/2$ ,  $x_n/2 \leq k_1 + \dots + k_{i-1} \leq x_n$  and  $x_n \leq k_1 + \dots + k_{i-1} \leq (i-1)(n \max_{j=1,\dots,i-1} p_j + \sqrt{n} \ln n) = (i-1)(n/m + \sqrt{n} \ln n)$ . When  $k_1 + \dots + k_{i-1} \leq x_n/2$ , (4.38) ensures that for all  $t > 0$ :

$$\begin{aligned} \Theta_{k_1+\dots+k_{i-1}}^l(x_n) &\leq \exp\left(t(k_1 + \dots + k_{i-1} - x_n) - (k_1 + \dots + k_{i-1}) \ln(2 - e^{-t})\right) \\ &\leq \exp\left(-\frac{t}{2} x_n\right). \end{aligned} \tag{4.41}$$

When  $x_n/2 \leq k_1 + \dots + k_{i-1} \leq x_n$ , (4.38) ensures that for all  $t > 0$ :

$$\begin{aligned} \Theta_{k_1+\dots+k_{i-1}}^l(x_n) &\leq \exp\left(t(k_1 + \dots + k_{i-1} - x_n) - (k_1 + \dots + k_{i-1}) \ln(2 - e^{-t})\right) \\ &\leq \exp\left(-\frac{x_n}{2} \ln(2 - e^{-t})\right). \end{aligned} \tag{4.42}$$

When  $x_n \leq k_1 + \dots + k_{i-1} \leq (i-1)(n/m + \sqrt{n} \ln n)$ , (4.39) and (4.33) in Lemma 4.4 ensure that:

$$\begin{aligned} \Theta_{k_1+\dots+k_{i-1}}^l(x_n) &\leq C \exp\left(-c(k_1 + \dots + k_{i-1}) \min\left(\frac{x_n}{k_1 + \dots + k_{i-1}}, \left(\frac{x_n}{k_1 + \dots + k_{i-1}}\right)^2\right)\right) \\ &= C \exp\left(-c \frac{x_n^2}{k_1 + \dots + k_{i-1}}\right) \end{aligned}$$

$$\leq C \exp \left( -c \frac{n(\ln n)^2}{(i-1)(n/m + \sqrt{n} \ln n)} \right). \quad (4.43)$$

Gathering together the bounds (4.41), (4.42) and (4.43) proves (4.40), for  $\bullet = l$ . Combining this last fact with the corresponding result for  $\bullet = r$ , and via (4.40) and (4.34), proves Proposition 4.1.  $\square$

Now, thanks to Proposition 4.1, to prove the convergence to zero, as  $n \rightarrow +\infty$ , of the first term on the right-hand side of (4.30), it is enough to prove the same result for

$$\mathbb{P} \left( \left\{ \max_{\mathbf{k} \in \mathcal{C}_{n,i}^\#} \left| \sum_{j=\ell_i+1}^{\ell_i+N_i^*-(k_1+\dots+k_{i-1})} Z_j^{(i)} \right| \geq \varepsilon \sqrt{n} \right\} \cap E_n^i \right), \quad (4.44)$$

where the  $Z_j^{(i)}$  are given in (4.5), i.e.,  $Z_j^{(i)} = (N_m^{T_i^{j-1}, T_i^j} - 1)/\sqrt{2}$ ,  $i = 1, \dots, m-1$ ,  $j \geq 1$ . Our next elementary proposition, the ultimate before closing this section, provides tail estimates on the partial sums of the  $Z_j$  (omitting the indices  $i$  for a while).

**Proposition 4.2** *Let  $(Z_j)_{j \geq 1}$  be iid random variables as in (4.5). Then, for suitable positive and finite constants  $c$  and  $C$ , all  $x > 0$ , and all positive integer  $k$ ,*

$$\mathbb{P} \left( \sum_{j=1}^k Z_j \geq x \right) \leq \min_{t>0} \left( \exp \left( - \left( t(x\sqrt{2} + k) + k \ln(2 - e^t) \right) \right) \right) =: \Theta_k^r(x\sqrt{2}), \quad (4.45)$$

$$\mathbb{P} \left( \sum_{j=1}^k Z_j \geq x \right) \leq C \exp \left( -c \min \left( \frac{x}{k}, \left( \frac{x}{k} \right)^2 \right) \right), \quad (4.46)$$

$$\mathbb{P} \left( \sum_{j=1}^k Z_j \leq -x \right) \leq \min_{t>0} \left( \exp \left( - \left( t(x\sqrt{2} - k) + k \ln(2 - e^{-t}) \right) \right) \right) =: \Theta_k^l(x\sqrt{2}), \quad (4.47)$$

$$\mathbb{P} \left( \sum_{j=1}^k Z_j \leq -x \right) \leq C \exp \left( -c \min \left( \frac{x}{k}, \left( \frac{x}{k} \right)^2 \right) \right), \quad \text{for } x \leq k. \quad (4.48)$$

**Proof.** Recall from (4.5) that  $Z_j = (N_m^{T_i^{j-1}, T_i^j} - 1)/\sqrt{2}$ ,  $i \neq m$ , and from (3.2),

$$\mathbb{E} \left[ x^{N_m^{T_i^{j-1}, T_i^j}} \right] = \frac{1}{2-x}. \quad (4.49)$$

Hence, using the notation in (4.36),

$$\mathbb{P} \left( \sum_{j=1}^k Z_j \geq x \right) \leq \min_{t>0} \left( e^{-t(x\sqrt{2}+k)} \mathbb{E} \left[ \exp \left( t N_m^{T_i^{j-1}, T_i^j} \right) \right]^k \right) = \Theta_k^r(x\sqrt{2}),$$

and (4.46) follows from (4.36) and (4.37) in (the proof of) Proposition 4.1 (with its notation) and from (4.33) in Lemma 4.4. Similarly, using the notation in (4.38)

$$\begin{aligned} \mathbb{P}\left(\sum_{j=1}^k Z_j \leq -x\right) &= \mathbb{P}\left(\sum_{j=1}^k \left(1 - N_m^{T_i^{j-1}, T_i^j}\right) \geq x\sqrt{2}\right) \\ &\leq \min_{t>0} \left( e^{-t(x\sqrt{2}-k)} \mathbb{E}\left[\exp\left(-tN_m^{T_i^{j-1}, T_i^j}\right)\right]^k \right) = \Theta_k^l(x\sqrt{2}). \end{aligned}$$

which is (4.47). As previously observed via (4.39), when  $x \leq k$ , the minimization for  $\Theta_k^l(x)$  occurs at  $t = \ln((2k-x)/(2k-2x))$ , and, once again, (4.33) on Lemma 4.4 ensure (4.48).  $\square$

We are now ready to move towards completing this section. From its very definition in (4.29),

$$\begin{aligned} \mathcal{C}_{n,i}^\# &= \mathcal{C}_n^{i-1} \cap \left\{ \mathbb{E}[N_i^*] \leq \ell_i = k_1 + \dots + k_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n - (N_i^* - \mathbb{E}[N_i^*]) \right\} \\ &\subset \left\{ k_1 + \dots + k_{i-1} \leq \sum_{j=1}^{i-1} \mathbb{E}[N_j] + (i-1)\sqrt{n} \ln n \right\} \\ &\quad \cap \left\{ \mathbb{E}[N_i^*] \leq \ell_i = k_1 + \dots + k_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n - (N_i^* - \mathbb{E}[N_i^*]) \right\} \\ &\subset \left\{ k_1 + \dots + k_{i-1} \leq (i-1)(n \max_{j=1, \dots, i-1} p_j + \sqrt{n} \ln n) \right\} \\ &\quad \cap \left\{ \mathbb{E}[N_i^*] \leq \ell_i = k_1 + \dots + k_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n - (N_i^* - \mathbb{E}[N_i^*]) \right\}. \end{aligned}$$

Therefore, recalling also from (3.25) that  $\mathbb{E}[N_i^*] = k_1 + \dots + k_{i-1}$ , (4.44) is upper bounded by:

$$\begin{aligned} &\mathbb{P}\left(\left\{ \max_{\substack{k_1 + \dots + k_{i-1} \leq (i-1)(n/m + \sqrt{n} \ln n) \\ k_1 + \dots + k_{i-1} \leq \ell_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n - (N_i^* - (k_1 + \dots + k_{i-1}))}} \left| \sum_{j=\ell_i+1}^{\ell_i + N_i^* - (k_1 + \dots + k_{i-1})} Z_j \right| \geq \varepsilon \sqrt{n} \right\} \cap E_n^i \right) \\ &\leq \mathbb{P}\left(\max_{\substack{k_1 + \dots + k_{i-1} \leq (i-1)(n/m + \sqrt{n} \ln n) \\ k_1 + \dots + k_{i-1} \leq \ell_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n + x_n}} \max_{|n_i| \leq x_n} \left| \sum_{j=\ell_i+1}^{\ell_i + n_i} Z_j \right| \geq \varepsilon \sqrt{n} \right) \quad (\text{recall (4.32)}) \\ &\leq \mathbb{P}\left(\max_{\ell_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n + x_n} \max_{|n_i| \leq x_n} \left| \sum_{j=\ell_i+1}^{\ell_i + n_i} Z_j \right| \geq \varepsilon \sqrt{n} \right) \\ &\leq 3nx_n \max_{\substack{\ell_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n + x_n \\ |n_i| \leq x_n}} \mathbb{P}\left(\left| \sum_{j=\ell_i+1}^{\ell_i + n_i} Z_j \right| \geq \varepsilon \sqrt{n} \right) \\ &\leq 3nx_n \max_{\substack{\ell_i \leq \mathbb{E}[N_i] + \sqrt{n} \ln n + x_n \\ 0 \leq n_i \leq x_n}} \left( \Theta_{n_i}^l(\varepsilon \sqrt{2n}) + \Theta_{n_i}^r(\varepsilon \sqrt{2n}) \right), \end{aligned} \tag{4.50}$$

where, in the next to last inequality, we used the usual (sharp in the iid case) bounding of the maximum via the number of terms times the maximal probability; while in the last one,  $|n_i| \leq x_n$  was changed into  $0 \leq n_i \leq x_n = \sqrt{n} \ln n$ .

Our final task is to show that

$$\lim_{n \rightarrow +\infty} nx_n \max_{0 \leq n_i \leq x_n} \Theta_{n_i}^\bullet(\varepsilon\sqrt{2n}) = 0, \quad (4.51)$$

for  $\bullet \in \{l, r\}$ . This relies again on Lemma 4.4 and Proposition 4.2. For  $\bullet = r$ , when  $k < \varepsilon\sqrt{2n}$ , (4.45) and (4.46) entail that,

$$\Theta_k^r(\varepsilon\sqrt{2n}) \leq C \exp(-c\varepsilon\sqrt{2n}); \quad (4.52)$$

while, for  $\varepsilon\sqrt{2n} \leq k \leq x_n$ , they entail that,

$$\Theta_k^r(\varepsilon\sqrt{2n}) \leq C \exp(-2c\varepsilon^2 n/k) \leq C \exp(-2c\varepsilon^2 n/x_n) = C \exp(-2c\varepsilon^2 \sqrt{n}/\ln n). \quad (4.53)$$

Therefore, for  $\bullet = r$ , (4.51) follows from (4.52) and (4.53). Let us now turn our attention to  $\bullet = l$ . When  $\varepsilon\sqrt{2n} \leq k \leq x_n$ , (4.48) entails that,

$$\Theta_k^l(\varepsilon\sqrt{2n}) \leq C \exp(-2c\varepsilon^2 n/k) \leq C \exp(-2c\varepsilon^2 n/x_n) = C \exp(-2c\varepsilon^2 \sqrt{n}/\ln n). \quad (4.54)$$

For  $k \leq \varepsilon\sqrt{n/2}$ , (4.47) entails that, for any  $t > 0$ ,

$$\Theta_k^l(\varepsilon\sqrt{2n}) \leq \exp\left(t(k - \varepsilon\sqrt{2n}) - k \ln(2 - e^{-t})\right) \leq \exp\left(-\varepsilon t \sqrt{n/2}\right). \quad (4.55)$$

For  $\varepsilon\sqrt{n/2} \leq k \leq \varepsilon\sqrt{2n}$ , (4.47) entails that, for any  $t > 0$ ,

$$\Theta_k^l(\varepsilon\sqrt{2n}) \leq \exp\left(t(k - \varepsilon\sqrt{n/2}) - k \ln(2 - e^{-t})\right) \leq \exp\left(-\varepsilon t \sqrt{n/2} \ln(2 - e^{-t})\right). \quad (4.56)$$

Therefore, for  $\bullet = l$ , (4.51) follows from (4.54), (4.55), and (4.56). Gathering all the intermediate results, for any  $i = 2, \dots, m-1$ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left( \left\{ \max_{\mathbf{k} \in \mathcal{C}_{n,i}^\#} \left| \sum_{j=\ell_i+1}^{\ell_i+N_i^*-(k_1+\dots+k_{i-1})} Z_j^{(i)} \right| \geq \varepsilon\sqrt{n} \right\} \cap E_n^i \right) = 0,$$

and therefore,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left( \max_{\mathbf{k} \in \mathcal{C}_n^*} \left| \sum_{j=k_1+\dots+k_{i-1}+1}^{N_i^*+k_i} \frac{Z_j^{(i)}}{\sqrt{n}} \right| \geq \varepsilon \right) = 0.$$

The goal of this section has thus been achieved: the quantities (4.18) and (4.19) have the same weak limit.

### 4.3 The Constraints

To deal with the third heuristic limit, we now need to obtain the convergence of the random set of constraints towards a deterministic set of constraints. This fact will follow from the various reductions obtained to date as well as new arguments developed from now on. To start with, let us recall two elementary facts about convergence in distribution.

The first fact asserts that if  $(f_n)_{1 \leq n \leq \infty}$  is a sequence of Borel functions such that  $x_n \rightarrow x_\infty$  implies that  $f_n(x_n) \rightarrow f_\infty(x_\infty)$ , and if  $(X_n)_{n \geq 1}$  is a sequence of random variables such that  $X_n \Rightarrow X_\infty$ , then  $f_n(X_n) \Rightarrow f_\infty(X_\infty)$ . Indeed, via the Skorohod representation theorem for  $C_0([0, 1])$ -valued random variables, there exists a probability space and  $C_0([0, 1])$ -valued random variables  $Y_n$ ,  $1 \leq n \leq \infty$ , such that  $Y_n \stackrel{\mathcal{L}}{=} X_n$ ,  $1 \leq n \leq \infty$ , and  $Y_n \rightarrow Y_\infty$  with probability one. But, by hypothesis,  $f_n(Y_n) \rightarrow f_\infty(Y_\infty)$ , with probability one. Therefore  $f_n(X_n) \Rightarrow f_\infty(X_\infty)$ .

The second elementary fact is as follows: Let  $(X_n)_{n \geq 1}$  be a sequence of random variables such that  $X_n^\pm \Rightarrow Y$ , then  $X_n \Rightarrow Y$ , where  $x^+ = \max(x, 0)$  and  $x^- = \min(x, 0)$ . Indeed,  $\mathbb{P}(X_n^+ \leq x) \leq \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X_n^- \leq x)$ , for all  $x \in \mathbb{R}$ .

Using these two elementary facts, let us return to our derandomization problem. Recalling (4.19), and using the polygonal structure of the processes  $B_n^X$  and  $B_n^Y$ , we have

$$M_n := \max_{\mathbf{k} \in \mathcal{C}_n} \left( F_X \left( B_n^X, \frac{\mathbf{k}}{n} \right) \wedge F_Y \left( B_n^Y, \frac{\mathbf{k}}{n} \right) \right),$$

where

$$F_X(\mathbf{u}, \mathbf{t}) = \frac{1}{m} \sum_{i=1}^{m-1} u_i(p_i(X)) - \sum_{i=1}^{m-1} \left( u_i \left( \sum_{j=1}^i t_j \right) - u_i \left( \sum_{j=1}^{i-1} t_j \right) \right), \quad (4.57)$$

$$F_Y(\mathbf{u}, \mathbf{t}) = \frac{1}{m} \sum_{i=1}^{m-1} u_i(p_i(Y)) - \sum_{i=1}^{m-1} \left( u_i \left( \sum_{j=1}^i t_j \right) - u_i \left( \sum_{j=1}^{i-1} t_j \right) \right), \quad (4.58)$$

for  $\mathbf{u} = (u_1, \dots, u_{m-1}) \in (C_0([0, 1]))^{m-1}$  and  $\mathbf{t} = (t_1, \dots, t_{m-1}) \in [0, 1]^{m-1}$ . Now, let

$$\mathcal{C}_n^\pm = \left\{ \mathbf{k} = (k_i)_{1 \leq i \leq m-1} : \forall i = 1, \dots, m-1, 0 \leq k_i \leq n \text{ and } \sum_{j=1}^i \frac{k_j}{n} \leq p_i \pm 2x_n \right\}, \quad (4.59)$$

with  $x_n = \sqrt{n} \ln(n)$  as in (4.31), and let

$$M_n^\pm = \max_{\mathbf{k} \in \mathcal{C}_n^\pm} \left( F_X \left( B_n^X, \frac{\mathbf{k}}{n} \right) \wedge F_Y \left( B_n^Y, \frac{\mathbf{k}}{n} \right) \right). \quad (4.60)$$

Since

$$N_i(X) - N_i^*(X) = np_i - \sum_{j=1}^{i-1} k_j + \left( (N_i(X) - \mathbb{E}[N_i(X)]) - (N_i^*(X) - \mathbb{E}[N_i^*(X)]) \right),$$



with a similar statement replacing  $X$  by  $Y$ , the condition

$$k_i \leq (N_i(X) - N_i^*(X)) \wedge (N_i(Y) - N_i^*(Y)),$$

in the definition (2.11)–(2.12) of  $\mathcal{C}_n$ , writes as  $\sum_{j=1}^i k_j/n \leq p_i + R_n^i(X, Y)/n$  where

$$\begin{aligned} R_n^i(X, Y) = & \left( (N_i(X) - \mathbb{E}[N_i(X)]) - (N_i^*(X) - \mathbb{E}[N_i^*(X)]) \right) \\ & \wedge \left( (N_i(Y) - \mathbb{E}[N_i(Y)]) - (N_i^*(Y) - \mathbb{E}[N_i^*(Y)]) \right). \end{aligned} \quad (4.61)$$

Now let

$$F_n := \bigcap_{i=1}^{m-1} \{ |N_i - \mathbb{E}[N_i]| \leq x_n \} \cap E_n^i,$$

with  $E_n^i$  defined in (4.32). From (4.10) and Proposition 4.1, we have  $\lim_{n \rightarrow +\infty} \mathbb{P}(F_n^c) = 0$  and, on  $F_n$ ,  $R_n^i(X, Y) \leq 2x_n$ , for all  $1 \leq i \leq m$ . Therefore, when  $F_n$  is realized,  $\mathcal{C}_n$  in (2.11) is encapsulated as follows:  $\mathcal{C}_n^- \subset \mathcal{C}_n \subset \mathcal{C}_n^+$ , and

$$M_n^- \leq M_n \leq M_n^+. \quad (4.62)$$

Clearly,

$$M_n^\pm = \max_{\mathbf{t} \in \mathcal{C}_n^\pm} (F_X(B_n^X, \mathbf{t}) \wedge F_Y(B_n^Y, \mathbf{t})), \quad (4.63)$$

where now

$$\mathcal{C}_n^\pm = \left\{ \mathbf{t} = (t_i)_{1 \leq i \leq m-1} \in [0, 1]^{m-1} : \forall i = 1, \dots, m-1, \sum_{j=1}^i t_j \leq p_i \pm 2 \frac{x_n}{n} \right\}. \quad (4.64)$$

Next,

$$\begin{aligned} \mathbb{P}(M_n \leq x) & \leq \mathbb{P}(\{M_n \leq x\} \cap F_n) + \mathbb{P}(F_n^c) \\ & \leq \mathbb{P}(\{M_n^- \leq x\} \cap F_n) + \mathbb{P}(F_n^c) \\ & \leq \mathbb{P}(M_n^- \leq x) + \mathbb{P}(F_n^c), \end{aligned}$$

therefore

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(M_n \leq x) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(M_n^- \leq x). \quad (4.65)$$

Similarly,

$$\begin{aligned} \mathbb{P}(M_n \leq x) & = \mathbb{P}(\{M_n \leq x\} \cap F_n) + \mathbb{P}(\{M_n \leq x\} \cap F_n^c) \\ & \geq \mathbb{P}(\{M_n^+ \leq x\} \cap F_n) \end{aligned}$$

$$\geq \mathbb{P}(M_n^+ \leq x) - \mathbb{P}(F_n^c),$$

and therefore

$$\liminf_{n \rightarrow +\infty} \mathbb{P}(M_n \leq x) \geq \liminf_{n \rightarrow +\infty} \mathbb{P}(M_n^+ \leq x). \quad (4.66)$$

Combining (4.65) and (4.66) with the second elementary fact described above, our goal is now to show that the convergence in distribution of both  $M_n^+$  and  $M_n^-$  towards

$$M_\infty = \max_{\mathbf{t} \in \mathcal{V}} (F_X(B^X, \mathbf{t}) \wedge F_Y(B^Y, \mathbf{t})), \quad (4.67)$$

holds true, where

$$\mathcal{V} := \mathcal{V}(p_1, \dots, p_{m-1}) = \left\{ \mathbf{t} = (t_j)_{1 \leq j \leq m-1} \in [0, 1]^{m-1} : \forall i = 1, \dots, m-1, \sum_{j=1}^i t_j \leq p_i \right\}.$$

To do so, first note that by Donsker's theorem  $(B_n^X, B_n^Y) \Rightarrow (B^X, B^Y)$  and we now wish to apply the first elementary fact, recalled above, to the functions

$$f_n^\pm(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{t} \in \mathcal{C}_n^\pm} (F_X(\mathbf{u}, \mathbf{t}) \wedge F_Y(\mathbf{v}, \mathbf{t})), \quad (4.68)$$

and

$$f_\infty(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{t} \in \mathcal{V}} (F_X(\mathbf{u}, \mathbf{t}) \wedge F_Y(\mathbf{v}, \mathbf{t})). \quad (4.69)$$

With these notations,  $M_n^\pm = f_n^\pm(B_n^X, B_n^Y)$  and  $M_\infty = f_\infty(B^X, B^Y)$ . In other words, we wish to show that  $(\mathbf{u}_n, \mathbf{v}_n) \rightarrow (\mathbf{u}, \mathbf{v})$  in  $(C_0([0, 1]))^{m-1}$  implies that  $f_n(\mathbf{u}_n, \mathbf{v}_n) \rightarrow f_\infty(\mathbf{u}, \mathbf{v})$ . To start with,

$$|f_n^\pm(\mathbf{u}_n, \mathbf{v}_n) - f_\infty(\mathbf{u}, \mathbf{v})| \leq |f_n^\pm(\mathbf{u}_n, \mathbf{v}_n) - f_n^\pm(\mathbf{u}, \mathbf{v})| + |f_n^\pm(\mathbf{u}, \mathbf{v}) - f_\infty(\mathbf{u}, \mathbf{v})|, \quad (4.70)$$

and we continue by estimating  $|f_n^\pm(\mathbf{u}_n, \mathbf{v}_n) - f_n^\pm(\mathbf{u}, \mathbf{v})|$ . But,

$$\begin{aligned} & |f_n^\pm(\mathbf{u}_n, \mathbf{v}_n) - f_n^\pm(\mathbf{u}, \mathbf{v})| \\ & \leq \max_{\mathbf{t} \in \mathcal{C}_n^\pm} \left| \left( F_X(\mathbf{u}_n, \mathbf{t}) \wedge F_Y(\mathbf{v}_n, \mathbf{t}) \right) - \left( F_X(\mathbf{u}, \mathbf{t}) \wedge F_Y(\mathbf{v}, \mathbf{t}) \right) \right| \\ & \leq \max_{\mathbf{t} \in \mathcal{C}_n^\pm} \max \left( |F_X(\mathbf{u}_n, \mathbf{t}) - F_X(\mathbf{u}, \mathbf{t})|, |F_Y(\mathbf{v}_n, \mathbf{t}) - F_Y(\mathbf{v}, \mathbf{t})| \right) \end{aligned} \quad (4.71)$$

$$\leq c \max_{\mathbf{t} \in \mathcal{C}_n^\pm} \max \left( |\mathbf{u}_n(\mathbf{t}) - \mathbf{u}(\mathbf{t})|, |\mathbf{v}_n(\mathbf{t}) - \mathbf{v}(\mathbf{t})| \right), \quad (4.72)$$

making use of Lemma 4.1 in (4.71), and by the linearity of both  $F_X$  and  $F_Y$ , with respect to their first argument, in (4.72) and where, further,  $c$  is a finite positive constant (depending explicitly on  $m$ ). Therefore,

$$|f_n^\pm(\mathbf{u}_n, \mathbf{v}_n) - f_n^\pm(\mathbf{u}, \mathbf{v})| \leq c \max(\|\mathbf{u}_n - \mathbf{u}\|_\infty, \|\mathbf{v}_n - \mathbf{v}\|_\infty),$$

and so if  $(\mathbf{u}_n, \mathbf{v}_n) \rightarrow (\mathbf{u}, \mathbf{v})$ , it follows that  $f_n^\pm(\mathbf{u}_n, \mathbf{v}_n) - f_n^\pm(\mathbf{u}, \mathbf{v}) \rightarrow 0$ .

In order to complete the proof of  $M_n^\pm \Rightarrow M_\infty$  and thus that of  $M_n \Rightarrow M_\infty$ , let us now estimate the right-most expression in (4.70).

At first, note that  $\mathcal{C}_n^- \subset \mathcal{V} \subset \mathcal{C}_n^+$ , hence

$$f_n^-(\mathbf{u}, \mathbf{v}) \leq f_\infty(\mathbf{u}, \mathbf{v}) \leq f_n^+(\mathbf{u}, \mathbf{v}). \quad (4.73)$$

Next, via (4.68) and (4.69), set  $f_n^+(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{t} \in \mathcal{C}_n^+} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t})$ , and  $f_\infty(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{t} \in \mathcal{V}} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t})$ , where  $\theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}) = F_X(\mathbf{u}, \mathbf{t}) \wedge F_Y(\mathbf{v}, \mathbf{t})$ . Since  $\mathcal{C}_n^- \subset \mathcal{C}_{n+1}^-$ , for  $n \geq 1$ , it follows (as shown next) that  $f_n^-(\mathbf{u}, \mathbf{v}) \rightarrow \max_{\mathbf{t} \in \bigcup_{n \geq 1} \mathcal{C}_n^-} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t})$ . Indeed,  $\lim_{n \rightarrow +\infty} f_n^-(\mathbf{u}, \mathbf{v}) \leq \max_{\mathbf{t} \in \bigcup_{n \geq 1} \mathcal{C}_n^-} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t})$  and if the previous inequality were strict, there would now be  $K \in (0, +\infty)$  such that

$$\max_{\mathbf{t} \in \mathcal{C}_n^-} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}) \leq K < \max_{\mathbf{t} \in \bigcup_{n \geq 1} \mathcal{C}_n^-} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}).$$

The left-hand side inequality implies that for all  $n \geq 1$ , and  $\mathbf{t} \in \mathcal{C}_n^-$ ,  $\theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}) \leq K$ , contradicting the right-hand side inequality.

Since  $\mathcal{C}_n^+ \supset \mathcal{C}_{n+1}^+$ , for  $n \geq 1$ , it also follows that  $f_n^+(\mathbf{u}, \mathbf{v}) \rightarrow \max_{\mathbf{t} \in \bigcap_{n \geq 1} \mathcal{C}_n^+} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t})$ . Indeed, we have  $\lim_{n \rightarrow +\infty} f_n^+(\mathbf{u}, \mathbf{v}) \geq \max_{\mathbf{t} \in \bigcap_{n \geq 1} \mathcal{C}_n^+} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t})$  and if the previous inequality were strict, there would be  $K \in (0, +\infty)$  such that

$$\max_{\mathbf{t} \in \mathcal{C}_n^+} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}) \geq K > \max_{\mathbf{t} \in \bigcap_{n \geq 1} \mathcal{C}_n^+} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}).$$

The left-hand side inequality implies that for any  $n \geq 1$ , there exists  $\mathbf{t}_n \in \mathcal{C}_n^+$  with  $\theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}_n) \geq K$ . Up to a subsequence  $\mathbf{t}_n \rightarrow \mathbf{t}^* \in \bigcap_{n \geq 1} \mathcal{C}_n^+$  and by the continuity of  $\theta_{\mathbf{u}, \mathbf{v}}$ ,  $\theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}^*) \geq K$ , which is inconsistent with the previous right-hand side inequality.

Finally, since  $\bigcup_{n \geq 1} \mathcal{C}_n^- = \mathcal{V}^\circ$ , the interior of  $\mathcal{V}$ , and since  $\bigcap_{n \geq 1} \mathcal{C}_n^+ = \overline{\mathcal{V}} = \mathcal{V}$ , the closure of  $\mathcal{V}$ , we have

$$\lim_{n \rightarrow +\infty} f_n^-(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{t} \in \mathcal{V}^\circ} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}) \leq f_\infty(\mathbf{u}, \mathbf{v}) = \max_{\mathbf{t} \in \mathcal{V}} \theta_{\mathbf{u}, \mathbf{v}}(\mathbf{t}) = \lim_{n \rightarrow +\infty} f_n^+(\mathbf{u}, \mathbf{v}). \quad (4.74)$$

It remains to show that the maximum of  $\theta_{\mathbf{u}, \mathbf{v}}$  on  $\mathcal{V}$  is attained on  $\mathcal{V}^\circ$  for  $\mathbb{P}_{(B^X, B^Y)}$ -almost all  $(\mathbf{u}, \mathbf{v})$ , i.e., that

$$\mathbb{P} \left( \max_{\mathbf{t} \in \mathcal{V}(1/m, \dots, 1/m)^\circ} \theta_{B^X, B^Y}(\mathbf{t}) = \max_{\mathbf{t} \in \mathcal{V}(1/m, \dots, 1/m)} \theta_{B^X, B^Y}(\mathbf{t}) \right) = 1. \quad (4.75)$$

With (4.75), (4.74) entails  $\lim_{n \rightarrow +\infty} f_n^\pm(\mathbf{u}, \mathbf{v}) = f_\infty(\mathbf{u}, \mathbf{v})$  for  $\mathbb{P}_{(B^X, B^Y)}$ -almost all  $(\mathbf{u}, \mathbf{v})$ , i.e., the right-most expression in (4.70) converges to 0 and, as previously explained, this gives  $M_n^\pm \Rightarrow M_\infty$  and  $M_n \Rightarrow M_\infty$ .

In order to complete (4.75) we anticipate, in the second equality below, on the results of Section 4.4 in which parameters are changed via:  $s_1 = u_1, s_1 + s_2 = u_2, \dots, s_1 + \dots + s_{m-1} = u_{m-1}$  and where we prove that

$$(\theta_{B^X, B^Y}(\mathbf{t}))_{\mathbf{t} \in \mathcal{V}(1/m, \dots, 1/m)} \stackrel{\mathcal{L}}{=} \frac{1}{\sqrt{m}} (\theta_{B^X, B^Y}(\mathbf{s}))_{\mathbf{s} \in \mathcal{V}(1, \dots, 1)} = \frac{1}{\sqrt{2m}} (\tilde{\theta}_{B_1, B_2}(\mathbf{u}))_{\mathbf{u} \in \mathcal{W}_m(1)},$$

where  $\mathcal{W}_m(1) = \{0 = u_0 \leq u_1 \leq \dots \leq u_{m-1} \leq u_m = 1\}$ ,

$$\begin{aligned} \tilde{\theta}_{B_1, B_2}(\mathbf{u}) = & \left( -\frac{1}{m} \sum_{i=1}^m B_1^{(i)}(1) + \sum_{i=1}^m \left( B_1^{(i)}(u_i) - B_1^{(i)}(u_{i-1}) \right) \right) \\ & \wedge \left( -\frac{1}{m} \sum_{i=1}^m B_2^{(i)}(1) + \sum_{i=1}^m \left( B_2^{(i)}(u_i) - B_2^{(i)}(u_{i-1}) \right) \right), \end{aligned} \quad (4.76)$$

and with  $B_1$  and  $B_2$  two independent, standard,  $m$ -dimensional Brownian on  $[0, 1]$ . The property (4.75) is thus equivalent to

$$\mathbb{P} \left( \max_{\mathbf{u} \in \mathcal{W}_m(1)^\circ} \tilde{\theta}_{B_1, B_2}(\mathbf{u}) = \max_{\mathbf{u} \in \mathcal{W}_m(1)} \tilde{\theta}_{B_1, B_2}(\mathbf{u}) \right) = 1. \quad (4.77)$$

The advantage of (4.77) over (4.75) is that the former involves two standard Brownian motions each one having *independent* coordinates. Roughly speaking, the property (4.77) should be derived from the following observation: when  $\mathbf{u} \in \partial \mathcal{W}_m(1)$ , then  $u_k = u_{k+1}$ , for some index  $k$ , and for such a  $\mathbf{u}$ , the sum  $\sum_{i=1}^m (B_1^{(i)}(u_i) - B_1^{(i)}(u_{i-1}))$  contains only  $m-1$  terms. Letting  $\mathbf{u}_\varepsilon$  be given by

$$u_{\varepsilon, i} = u_i, \quad i \neq k+1, \quad \text{and} \quad u_{\varepsilon, k+1} = u_k + \varepsilon,$$

we have

$$\begin{aligned} & \sum_{i=1}^m (B_1^{(i)}(u_{\varepsilon, i}) - B_1^{(i)}(u_{\varepsilon, i-1})) \\ &= \sum_{i=1}^m (B_1^{(i)}(u_i) - B_1^{(i)}(u_{i-1})) + (B_1^{(k+1)}(u_k + \varepsilon) - B_1^{(k+1)}(u_k)) \\ & \quad + (B_1^{(k+2)}(u_k) - B_1^{(k+2)}(u_k + \varepsilon)). \end{aligned}$$

The terms  $(B_1^{(k+1)}(u_k + \varepsilon) - B_1^{(k+1)}(u_k))$  and  $(B_1^{(k+2)}(u_k) - B_1^{(k+2)}(u_k + \varepsilon))$  are independent of  $\sum_{i=1}^m (B_1^{(i)}(u_i) - B_1^{(i)}(u_{i-1}))$  and from standard properties of Brownian motion, almost surely, the sum  $(B_1^{(k+1)}(u_k + \varepsilon) - B_1^{(k+1)}(u_k)) + (B_1^{(k+2)}(u_k) - B_1^{(k+2)}(u_k + \varepsilon))$  takes positive value for arbitrarily small  $\varepsilon > 0$ . Since the same is true for the second term in (4.76) relative to  $B_2$ , it follows that in the vicinity of each  $\mathbf{u} \in \partial \mathcal{W}_m(1)$ , there is  $\mathbf{u}_\varepsilon \in \mathcal{W}_m(1)$  with  $\tilde{\theta}_{B_1, B_2}(\mathbf{u}_\varepsilon) > \tilde{\theta}_{B_1, B_2}(\mathbf{u})$ . Therefore,  $\max_{\mathbf{u} \in \mathcal{W}_m(1)} \tilde{\theta}_{B_1, B_2}(\mathbf{u})$  is attained in  $\mathcal{W}_m(1)^\circ$ , and so both (4.77) and (4.75) hold true, leading to  $M_n \Rightarrow M_\infty$ .

#### 4.4 Final Step: A Linear Transformation

By combining the results of the previous three subsections, we proved that

$$\frac{\text{LCI}_n - n/m}{\sqrt{2n}} \Rightarrow \max_{\nu(1/m, \dots, 1/m)} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B^{(i), X} \left( \frac{1}{m} \right) - \sum_{i=1}^{m-1} \left( B^{(i), X} \left( \sum_{j=1}^i t_j \right) - B^{(i), X} \left( \sum_{j=1}^{i-1} t_j \right) \right) \right),$$

$$\frac{1}{m} \sum_{i=1}^{m-1} B^{(i),Y} \left( \frac{1}{m} \right) - \sum_{i=1}^{m-1} \left( B^{(i),Y} \left( \sum_{j=1}^i t_j \right) - B^{(i),Y} \left( \sum_{j=1}^{i-1} t_j \right) \right) \quad (4.78)$$

where the maximum is taken over  $\mathbf{t} = (t_1, \dots, t_{m-1}) \in \mathcal{V}(1/m, \dots, 1/m)$ . Now, via the linear transformations of the parameters given by  $s_i = m \sum_{j=1}^i t_j$ ,  $i = 1, \dots, m-1$ ,  $s_0 = t_0 = 0$ , and Brownian scaling, the right-hand side of (4.78) becomes equal, in law, to:

$$\frac{1}{\sqrt{m}} \max_{0=s_0 \leq s_1 \leq \dots \leq s_{m-1} \leq 1} \min \left( \frac{1}{m} \sum_{i=1}^{m-1} B^{(i),X}(1) - \sum_{i=1}^{m-1} (B^{(i),X}(s_i) - B^{(i),X}(s_{i-1})), \right. \\ \left. \frac{1}{m} \sum_{i=1}^{m-1} B^{(i),Y}(1) - \sum_{i=1}^{m-1} (B^{(i),Y}(s_i) - B^{(i),Y}(s_{i-1})) \right). \quad (4.79)$$

Next, for all  $t \in [0, 1]$  and  $i = 1, \dots, m-1$ , let us introduce the following two pointwise linear transformations:

$$B^{(i),X}(t) = \frac{B_1^{(m)}(t) - B_1^{(i)}(t)}{\sqrt{2}}, \\ B^{(i),Y}(t) = \frac{B_2^{(m)}(t) - B_2^{(i)}(t)}{\sqrt{2}},$$

where  $B_1$  and  $B_2$  are two, standard,  $m$ -dimensional Brownian motion on  $[0, 1]$ . Clearly  $(B^{(1),X}(t), \dots, B^{(m-1),X}(t))_{0 \leq t \leq 1}$  has the correct covariance matrix (4.13), and similarly for  $B_2$ , replacing  $X$  by  $Y$ . Moreover,

$$\frac{1}{m} \sum_{i=1}^{m-1} B^{(i),X}(1) - \sum_{i=1}^{m-1} (B^{(i),X}(s_i) - B^{(i),X}(s_{i-1})) \\ = -\frac{1}{\sqrt{2}m} \left( \sum_{i=1}^m B_1^{(i)}(1) \right) + \frac{1}{\sqrt{2}} B_1^{(m)}(1) \\ - \frac{1}{\sqrt{2}} \sum_{i=1}^{m-1} (B_1^{(m)}(s_i) - B_1^{(m)}(s_{i-1})) + \frac{1}{\sqrt{2}} \sum_{i=1}^{m-1} (B_1^{(i)}(s_i) - B_1^{(i)}(s_{i-1})) \\ = \frac{1}{\sqrt{2}} \left( -\frac{1}{m} \sum_{i=1}^m B_1^{(i)}(1) + (B_1^{(m)}(1) - B_1^{(m)}(s_{m-1})) + \sum_{i=1}^{m-1} (B_1^{(i)}(s_i) - B_1^{(i)}(s_{i-1})) \right). \quad (4.80)$$

Finally, with the help of (4.80) (and the corresponding identity for  $Y$ ), (4.79) becomes:

$$\frac{1}{\sqrt{2}m} \max_{0=s_0 \leq s_1 \leq \dots \leq s_{m-1} \leq s_m=1} \min \left( -\frac{1}{m} \sum_{i=1}^m B_1^{(i)}(1) + \sum_{i=1}^m (B_1^{(i)}(s_i) - B_1^{(i)}(s_{i-1})), \right. \\ \left. -\frac{1}{m} \sum_{i=1}^m B_2^{(i)}(1) + \sum_{i=1}^m (B_2^{(i)}(s_i) - B_2^{(i)}(s_{i-1})) \right), \quad (4.81)$$

and the proof of Theorem 1.1 is over.

## 5 Concluding Remarks

Let us discuss below some potential extensions to Theorem 1.1 and some questions we believe are of interest.

- From the proof presented above, the passage from two to three or more sequences is clear: the minimum over two Brownian functionals becomes a minimum over three or more Brownian functionals, and such a passage applies to the cases touched upon below.

- It is also clear from the proof developed above, that a theorem for two sequences of iid (non-uniform) random variables is also valid. Here is what it should look like: Let  $X = (X_i)_{i \geq 1}$  and  $Y = (Y_i)_{i \geq 1}$  be two sequences of iid random variables with values in  $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < \dots < \alpha_m\}$ , a totally ordered finite alphabet of cardinality  $m$  and with a common law, i.e.,  $X_1 \stackrel{\mathcal{L}}{=} Y_1$ . Let  $p_{\max} = \max_{i=1,2,\dots,m} \mathbb{P}(X_1 = \alpha_i)$  and let  $k$  be the multiplicity of  $p_{\max}$ . Then,

$$\frac{\text{LCI}_n - np_{\max}}{\sqrt{np_{\max}}} \Rightarrow \max_{0=t_0 \leq t_1 \leq \dots \leq t_{k-1} \leq t_k=1} \min \left( \frac{\sqrt{1 - kp_{\max}} - 1}{k} \sum_{i=1}^k B_1^{(i)}(1) + \sum_{i=1}^k (B_1^{(i)}(t_i) - B_1^{(i)}(t_{i-1})), \right. \\ \left. \frac{\sqrt{1 - kp_{\max}} - 1}{k} \sum_{i=1}^k B_2^{(i)}(1) + \sum_{i=1}^k (B_2^{(i)}(t_i) - B_2^{(i)}(t_{i-1})) \right), \quad (5.1)$$

where  $B_1$  and  $B_2$  are two  $k$ -dimensional standard Brownian motions defined on  $[0, 1]$ . So, for instance, if  $p_{\max}$  is uniquely attained then the limiting law in (5.1) is the minimum of two centered Gaussian random variables.

Using the sandwiching techniques developed in [HL], an infinite countable alphabet result can also be obtained with (5.1).

- The loss of independence inside the sequences, and the loss of identical distributions, both within and between the sequences is more challenging. Results for these situations will be presented elsewhere.

- The length of the longest increasing subsequence of a random word is well known to have an equivalent interpretation in percolation theory: Indeed, consider the following directed last-passage percolation model in  $\mathbb{Z}_+^2$ : let  $\Pi_2(n, m)$  be the set of directed paths in  $\mathbb{Z}_+^2$  from  $(0, 0)$  to  $(n, m)$  with unit steps going either North or East. Given random variables  $\omega_{i,j}$ ,  $i \geq 0, j \geq 1$ , and interpreting each  $\omega_{i,j}$  as the length of time spent by a path at the vertex  $(i, j)$ , the last-passage time to  $(n, m)$  is given by

$$T_2(n, m) = \max_{\pi \in \Pi_2(n, m)} \left( \sum_{(i,j) \in \pi} \omega_{i,j} \right). \quad (5.2)$$

(See Bodineau and Martin [BM], and the references therein, for details.) In our random word context, when  $X = (X_i)_{1 \leq i \leq n}$  is a sequence of iid random variables taking their values

in a totally ordered finite alphabet  $\{\alpha_1 < \alpha_2 < \dots < \alpha_m\}$  of size  $m$ , taking  $\omega_{i,j} = \mathbf{1}_{\{X_i = \alpha_j\}}$  and  $\omega_{0,j} = 0$ ,  $j \geq 1$ , which for each  $i$  are *dependent* random variables, the length of the longest increasing subsequence of the random word is equal to the last passage-time  $T_2(n, m)$ , see [BH].

Now  $\text{LCI}_n$ , the length of the longest *common* and increasing subsequences, enjoys a similar percolation theory interpretation, but in  $\mathbb{Z}_+^3$ . Let  $\Pi_3(n, n, m)$  be the set of paths in  $\mathbb{Z}_+^3$  from  $(0, 0, 0)$  to  $(n, n, m)$  taking either unit steps towards the top or steps, of any length, in the horizontal plane but neither parallel to the  $x$ -axis nor to the  $y$ -axis, i.e.,

$$\Pi_3(n, n, m) := \left\{ (u_1, u_2, \dots, u_{n+m}) \in (\mathbb{Z}_+^3)^{n+m} : u_1 = (0, 0, 1), u_{n+m} = (n, n, m), \right. \\ \left. u_{j+1} - u_j \in \{(0, 0, 1), (a, b, 0) \text{ with } a, b \in \mathbb{N} \setminus \{0\}\}, j = 1, \dots, n+m-1 \right\}.$$

Given weights  $\omega_{i,j,k}$ ,  $i \geq 0, j \geq 0, k \geq 1$ , on the lattice, we can consider a quantity analogous to  $T_2(n, m)$  in (5.2), namely,

$$T_3(n, n, m) := \max_{\pi \in \Pi_3(n, n, m)} \left( \sum_{(i,j,k) \in \pi} \omega_{i,j,k} \right).$$

In the random word context, taking  $\omega_{i,j,k} = \mathbf{1}_{\{X_i = Y_j = \alpha_k\}}$  and  $\omega_{0,0,k} = 0$ ,  $k \geq 1$ , as weights, gives  $\text{LCI}_n = T_3(n, n, m)$ .

Note that when  $X = Y$ ,  $T_3(n, n, m)$  recovers  $T_2(n, m)$  since  $T_2(n, m)$  is unchanged if, in (5.2),  $\Pi_2(n, m)$  is replaced by

$$\tilde{\Pi}_2(n, m) := \left\{ (u_1, u_2, \dots, u_{n+m}) \in (\mathbb{Z}_+^2)^{n+m} : u_1 = (0, 1), u_{n+m} = (n, m), \right. \\ \left. u_{j+1} - u_j \in \{(0, 1), (a, b) \text{ with } a, b \in \mathbb{N} \setminus \{0\}\}, j = 1, \dots, n+m-1 \right\}.$$

More generally, for  $p \geq 3$  sequences of letters  $X^{(\ell)} = (X_i^{(\ell)})_{1 \leq i \leq n}$ ,  $1 \leq \ell \leq p$ , we can similarly consider

$$\Pi_{p+1}(n, \dots, m) := \left\{ (u_1, u_2, \dots, u_{n+m}) \in (\mathbb{Z}_+^{p+1})^{n+m} : u_1 = (0, \dots, 0, 1), u_{n+m} = (n, \dots, n, m) \right. \\ \left. u_{j+1} - u_j \in \{(0, \dots, 0, 1), (a_1, \dots, a_p, 0) \text{ with } a_i \in \mathbb{N} \setminus \{0\}\}, j = 1, \dots, n+m-1 \right\},$$

and

$$T_p(n, \dots, n, m) := \max_{\pi \in \Pi_{p+1}(n, \dots, n, m)} \left( \sum_{(i_1, \dots, i_p, k) \in \pi} \omega_{i_1, \dots, i_p, k} \right).$$

Then, observe that  $\text{LCI}_n$ , for the  $p$  sequences, is equal to  $T_p(n, \dots, n, m)$ , where now  $\omega_{i_1, \dots, i_p, k} = \mathbf{1}_{\{X_{i_1} = \dots = X_{i_p} = \alpha_k\}}$  and  $\omega_{0, \dots, 0, k} = 0$ ,  $k = 1, \dots, m$ , are dependent random variables.

In view of Theorem 1.1 and of [BM], one would expect that for  $m$  fixed and for exponential mean one iid weights  $\omega_{\dots}$ ,  $T_3(n, n, m)$  converges, when properly centered, by  $n$ , and scaled, by  $\sqrt{n}$ , towards

$$\max_{0=t_0 \leq t_1 \leq \dots \leq t_{m-1} \leq t_m=1} \min \left( \sum_{i=1}^m \left( B_1^{(i)}(t_i) - B_1^{(i)}(t_{i-1}) \right), \sum_{i=1}^m \left( B_2^{(i)}(t_i) - B_2^{(i)}(t_{i-1}) \right) \right),$$

with also the trivial modification for  $T_p$ .

- Starting with Baryshnikov [Bar] and Gravner, Tracy and Widom [GTW] (see, also [BGH], for a further description and up to date references) a strong interaction has been shown to exist between Brownian functionals, originating in queuing theory with Glynn and Whitt [GW] (see also Seppäläinen [Sep]), and maximal eigenvalues of Gaussian random matrices. Likewise, we hypothesize that the max/min functionals obtained here do enjoy a similar strong connection (which might extend to spectra and Young diagrams). Could it be that the right-hand side of (1.1) (with or without the linear terms) has the same law as the maximal eigenvalue of a random matrix model? Even in the binary case, it would be interesting to find the law of the processes  $(\sqrt{2} \max_{0 \leq t \leq 1} \min(B_1(t) - B_1(1)/2, B_2(t) - B_2(1)/2))_{t \geq 0}$  and  $(\max_{0 \leq t \leq 1} \min(B_1(t), B_2(t)))_{t \geq 0}$  where, say,  $B_1$  and  $B_2$  are two independent standard linear Brownian motions. Very preliminary work on these problems was started with Marc Yor, before his untimely death, and this text is dedicated to his memory.

- To finish, note that the LCIS problem for two or more uniform random permutations of  $\{1, 2, \dots, n\}$  has not been studied either, although it certainly deserves to be. In point of fact, it is shown in [HI] that, for any two independent uniform random permutations  $\sigma_1$  and  $\sigma_2$  of  $\{1, 2, \dots, n\}$ , and for any  $x \in \mathbb{R}$ ,  $\mathbb{P}(LC_n(\sigma_1, \sigma_2) \leq x) = \mathbb{P}(LI_n(\sigma_1) \leq x)$ , where  $LI_n(\sigma_1)$  is the length of the longest increasing subsequences of  $\sigma_1$ . Therefore, this equality in law shows the emergence of the Tracy-Widom distribution, which had sometimes been speculated, as the corresponding limiting law. Indeed, once we are given the result of Baik, Deift and Johansson [BDJ] on the limiting law of  $LI_n(\sigma_1)$ , a corresponding result (actually equivalent to it) for  $LC_n(\sigma_1, \sigma_2)$  is immediate. In fact, many of the results on  $LI_n(\sigma_1)$  presented in Romik [Rom], such as the law of large numbers of Vershik and Kerov [VK] are instantaneously transferable to equivalent versions for  $LC_n(\sigma_1, \sigma_2)$ .

Moreover, for  $p \geq 3$  independent and uniform random permutations  $\sigma_1, \sigma_2, \dots, \sigma_p$ , the methodology developed in [HI] easily shows that  $LC_n(\sigma_1, \sigma_2, \dots, \sigma_p) \stackrel{d}{=} LCIn(\sigma_1, \dots, \sigma_{p-1})$ , where  $\stackrel{d}{=}$  denotes equality in distribution. Therefore, the study of longest common and increasing subsequences in random words or random permutations which might appear, at first, quite artificial is actually intimately related to the study of longest common subsequences.

## A Appendix

### A.1 Proofs of technical lemmas

#### Proof of Lemma 4.1

First,

$$\begin{aligned} & \left| \max_{k=1, \dots, K} (a_k \wedge b_k) - \max_{k=1, \dots, K} ((a_k + c_k) \wedge (b_k + d_k)) \right| \\ & \leq \max_{k=1, \dots, K} |(a_k \wedge b_k) - ((a_k + c_k) \wedge (b_k + d_k))|. \end{aligned}$$



Next, the result will follow from the elementary inequality

$$(a \wedge b) - (a + c) \wedge (b + d) \leq |c| \vee |d|, \quad (\text{A.1})$$

which is valid for all  $a, b, c, d \in \mathbb{R}$ . Indeed, set  $D = (a \wedge b) - (a + c) \wedge (b + d)$  and assume (without loss of generality) that  $a \leq b$ . If  $a + c \leq b + d$ , then  $D = a - (a + c) = -c \leq |c|$ . If  $b + d \leq a + c$ , then  $D = a - b - d$  and so whenever  $a \leq b + d$ , (A.1) is immediate, while if  $a \geq b + d$ , then  $D = a - b - d \leq -d = |d|$  since  $a - b \leq 0$  and  $-d \geq b - a \geq 0$ .  $\square$

### Proof of Lemma 4.2

Let  $D_n = \{ |N^{(n)} - \mathbb{E}[N^{(n)}]| < x_n \}$ , and for  $\varepsilon > 0$ , let

$$A_n(\varepsilon) = \left\{ \left| \sum_{j \in [N^{(n)}, \mathbb{E}[N^{(n)}]]} \frac{Z_j}{\sqrt{n}} \right| \geq \varepsilon \right\}.$$

Since  $\mathbb{P}(A_n(\varepsilon)) \leq \mathbb{P}(A_n(\varepsilon) \cap D_n) + \mathbb{P}(D_n^c)$ , and since  $\lim_{n \rightarrow \infty} \mathbb{P}(D_n^c) = 0$ , and it is enough to show  $\lim_{n \rightarrow +\infty} \mathbb{P}(A_n(\varepsilon) \cap D_n) = 0$ . But, by Kolmogorov's maximal inequality,

$$\begin{aligned} \mathbb{P}(A_n(\varepsilon) \cap D_n) &\leq \mathbb{P} \left( \max_{|k - \mathbb{E}[N^{(n)}]| < x_n} \left| \sum_{j \in [k, \mathbb{E}[N^{(n)}]]} \frac{Z_j}{\sqrt{n}} \right| \geq \varepsilon \right) \\ &\leq \frac{x_n \text{Var}(Z_1)}{\varepsilon^2 n} \rightarrow 0, \quad n \rightarrow +\infty. \end{aligned}$$

$\square$

### Proof of Lemma 4.3

First, we show that  $(B_n^{(k)}(N_k/n)^2)_{n \geq 1}$  is uniformly integrable. Proposition 3.2 and Remark 3.1 give

$$\begin{aligned} B_n^{(k)}\left(\frac{N_k}{n}\right) &= \frac{N_m - N_k}{\sqrt{2n}} + o_{\mathbb{P}}(1/\sqrt{n}) \\ \mathbb{E} \left[ \left| B_n^{(k)}\left(\frac{N_k}{n}\right) \right|^p \right] &\leq 2^{p-1} \left( \mathbb{E} \left[ \left| \frac{N_m - N_k}{\sqrt{2n}} \right|^p \right] + o(n^{-p/2}) \right) \\ &= 2^{-1/2} n^{-p/2} \mathbb{E} \left[ |N_m - N_k|^p \right] + o(n^{-p/2}). \end{aligned}$$

But  $N_m - N_k = \sum_{i=1}^n \epsilon_i^{(m,k)}$  where  $(\epsilon_i^{(m,k)})_{i \geq 1}$  are iid with  $\epsilon_i^{(m,k)} = 1$  when  $X_i = \alpha_m$ ,  $\epsilon_i^{(m,k)} = -1$  when  $X_i = \alpha_k$  and  $\epsilon_i^{(m,k)} = 0$  otherwise. Hence, by the classical Marcinkiewicz-Zygmund inequality, for some constant  $C_p$ ,

$$\mathbb{E} \left[ |N_m - N_k|^p \right] = \mathbb{E} \left[ \left| \sum_{i=1}^n \epsilon_i \right|^p \right]$$

$$\begin{aligned}
&\leq C_p \mathbb{E} \left[ \left( \sum_{i=1}^n |\epsilon_i^{(m,k)}|^2 \right)^{p/2} \right] \\
&\leq C_p n^{p/2}.
\end{aligned}$$

Therefore, for any  $p > 2$

$$\sup_{n \geq 1} \mathbb{E} \left[ \left| B_n^{(k)} \left( \frac{N_k}{n} \right) \right|^p \right] < +\infty$$

and  $(B_n^{(k)} (N_k/n)^2)_{n \geq 1}$  is uniformly integrable. Next, for  $(B_n^{(k)} (1/m))_{n \geq 1}$ :

$$B_n^{(k)} \left( \frac{1}{m} \right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[n/m]} Z_j^{(k)} + \frac{(nm - [nm])}{\sqrt{n}} Z_{[n/m]+1}^{(k)}.$$

and

$$\begin{aligned}
\mathbb{E} \left[ \left| B_n^{(k)} \left( \frac{1}{m} \right) \right|^p \right] &\leq 2^{p-1} n^{-p/2} \mathbb{E} \left[ \left| \sum_{j=1}^{[n/m]} Z_j^{(k)} \right|^p \right] + 2^{p-1} n^{-p/2} \mathbb{E} \left[ \left| Z_{[n/m]+1}^{(k)} \right|^p \right] \\
&\leq C_p n^{-p/2} \mathbb{E} \left[ \left( \sum_{j=1}^{[n/m]} |Z_j^{(k)}|^2 \right)^{p/2} \right] + 2^{p-1} n^{-p/2} \mathbb{E} \left[ \left| Z_1^{(k)} \right|^p \right],
\end{aligned}$$

using again the Marcinkiewicz-Zygmund inequality. Continuing, using convexity,

$$C_p n^{-p/2} \mathbb{E} \left[ \left( \sum_{j=1}^{[n/m]} |Z_j^{(k)}|^2 \right)^{p/2} \right] \leq C_p n^{-p/2} \mathbb{E} \left[ [n/m]^{p/2-1} \sum_{j=1}^{[n/m]} |Z_j^{(k)}|^p \right] \leq \frac{C_p}{m^{p/2}} \mathbb{E} \left[ |Z_1^{(k)}|^p \right].$$

Hence, for any  $p > 2$ ,

$$\sup_{n \geq 1} \mathbb{E} \left[ \left| B_n^{(k)} \left( \frac{1}{m} \right) \right|^p \right] < +\infty,$$

and  $(B_n^{(k)} (1/m)^2)_{n \geq 1}$  is uniformly integrable and therefore, from above, so is  $(B_n^{(k)} (N_k/n) - B_n^{(k)} (1/m))^2$ . Finally, in order to show (4.17), it is enough to prove

$$B_n^{(k)} \left( \frac{N_k}{n} \right) - B_n^{(k)} \left( \frac{1}{m} \right) \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow +\infty. \quad (\text{A.2})$$

Setting  $A_n = \{|N_k - n/m| \leq \sqrt{n} \ln n\}$ , Hoeffding's inequality ensures that  $\lim_{n \rightarrow +\infty} \mathbb{P}(A_n^c) = 0$ . Therefore, since

$$B_n^{(k)} \left( \frac{N_k}{n} \right) - B_n^{(k)} \left( \frac{1}{m} \right) = \frac{1}{\sqrt{n}} \sum_{j=[n/m]+1}^{N_k} Z_j^{(k)},$$

we have

$$\mathbb{P} \left( \left| B_n^{(k)} \left( \frac{N_k}{n} \right) - B_n^{(k)} \left( \frac{1}{m} \right) \right| \geq \varepsilon \right) \leq \mathbb{P} \left( \left\{ \left| \sum_{j=[n/m]+1}^{N_k} Z_j^{(k)} \right| \geq \varepsilon \sqrt{n} \right\} \cap A_n \right) + \mathbb{P}(A_n^c),$$

and Kolmogorov's maximal inequality entails that

$$\begin{aligned} \mathbb{P}\left(\max_{l \in [n/m - \sqrt{n} \ln n, n/m + \sqrt{n} \ln n]} \left| \sum_{j=[n/m]+1}^l Z_j^{(k)} \right| \geq \varepsilon \sqrt{n}\right) &\leq \frac{1}{\varepsilon^2 n} \mathbb{E} \left[ \sum_{j=[n/m]+1}^{n/m + \sqrt{n} \ln n} (Z_j^{(k)})^2 \right] \\ &\leq \frac{(\ln n) \mathbb{E}[(Z_1^{(k)})^2]}{\sqrt{n}}, \end{aligned}$$

finishing the proof of (A.2) and thus of (4.17).

#### Proof of Lemma 4.4

Consider three cases:  $|x| \ll n$ ,  $x \gg n$  (here  $u_n \ll v_n$  means  $\lim_{n \rightarrow +\infty} u_n/v_n = 0$ ) and  $x \approx n$ , i.e.,  $c_1 n \leq x \leq c_2 n$ , for two finite constants  $c_1$  and  $c_2$ , and expand  $K_n(x)$  accordingly. First, let  $|x| \ll n$ : then,

$$\begin{aligned} K_n(x) &= \frac{(2n)^{x+2n}}{(2n)^{x+n}(2n)^n} \frac{(1 + \frac{x}{2n})^{x+2n}}{(1 + \frac{x}{n})^{x+n}} \\ &= \exp \left( (x+2n) \ln \left( 1 + \frac{x}{2n} \right) - (x+n) \ln \left( 1 + \frac{x}{n} \right) \right) \\ &= \exp \left( (x+2n) \left( \frac{x}{2n} - \frac{x^2}{8n^2} + o\left(\frac{x^2}{n^2}\right) \right) - (x+n) \left( \frac{x}{n} - \frac{x^2}{2n^2} + o\left(\frac{x^2}{n^2}\right) \right) \right) \\ &= \exp \left( -\frac{x^2}{4n} + \frac{3x^3}{8n^2} + o\left(\frac{x^3}{n^2}\right) + o\left(\frac{x^2}{n}\right) \right) \\ &= \exp \left( -\frac{x^2}{4n} + o\left(\frac{x^2}{n}\right) \right), \end{aligned}$$

which yields (4.33) in case  $|x| \ll n$ . Next, let  $x \gg n$ : then,

$$\begin{aligned} K_n(x) &= \frac{(x+2n)^{x+2n}}{(2x+2n)^{x+n}(2n)^n} = \frac{x^n}{(4n)^n 2^x} \frac{(1 + \frac{2n}{x})^{x+2n}}{(1 + \frac{n}{x})^{x+n}} \\ &= \frac{x^n}{(4n)^n 2^x} \exp \left( (x+2n) \ln \left( 1 + \frac{2n}{x} \right) - (x+n) \ln \left( 1 + \frac{n}{x} \right) \right) \\ &= \frac{x^n}{(4n)^n 2^x} \exp \left( (x+2n) \left( \frac{2n}{x} - \frac{2n^2}{x^2} + o\left(\frac{n^2}{x^2}\right) \right) - (x+n) \left( \frac{n}{x} - \frac{n^2}{2x^2} + o\left(\frac{n^2}{x^2}\right) \right) \right) \\ &= \frac{x^n}{(4n)^n 2^x} \exp \left( n + \frac{3n^2}{2x} - \frac{7n^3}{2x^2} + o\left(\frac{n^2}{x}\right) \right) \\ &= \exp \left( n + \frac{3n^2}{2x} + n \ln \left( \frac{x}{4n} \right) - x \ln 2 + o\left(\frac{n^2}{x}\right) \right). \end{aligned} \tag{A.3}$$

Since  $x \gg n$ , the larger order in the exponential (A.3) is  $x \ln 2$  and, this recover a bound of the form (4.33) in this case. Finally, consider the case  $x \approx n$ , say  $x = \alpha n$  with  $\alpha > -1$ .

Then,

$$K_n(x) = \frac{((\alpha + 2)n)^{(\alpha+2)n}}{((2\alpha + 2)n)^{(\alpha+1)n} (2n)^n} = \exp(-c(\alpha)n),$$

which is again of the form (4.33), since  $c(\alpha) = \ln(2(2\alpha + 2)^{\alpha+1}/(\alpha + 2)^{\alpha+2})$  is positive for all  $\alpha > -1$  and is also bounded.  $\square$

## A.2 On [HLM]

The purpose of this Appendix is to provide some missing steps in the proof of the main theorem in [HLM] devoted to the binary case as well as to correct the errors present there. The notations and numbering are as in [HLM]. In particular, recall that  $N_1$  (resp.  $N_2$ ) is the number of zeros in  $X_1, \dots, X_n$  (resp.  $Y_1, \dots, Y_n$ ).

**Proof of (13).** Recall again from [HLM] that

$$\begin{aligned} V_n &= \max_{0 \leq k \leq N_1 \wedge N_2} \left( \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)} \left( \frac{k}{n} \right) \right) \right), \\ X_n &= \max_{0 \leq t \leq \frac{1}{2}} \left( \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)}(t) \right) \right). \end{aligned}$$

Clearly,

$$X_n \geq \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) \right) = \frac{1}{2} \bigwedge_{i=1,2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right), \quad (\text{A.4})$$

and denote by  $i_*$  the index for which the minimum in (A.4) is attained.

Next, if  $N_1 \wedge N_2 \leq n/2$ , then  $V_n \leq X_n$ ; and similarly if the maximum defining  $V_n$  is attained at some  $k^* \leq n/2$ , then  $V_n \leq X_n$ . Otherwise,  $N_1 \wedge N_2 \geq n/2$  with, moreover, the maximum defining  $V_n$  attained at  $k^* \in [n/2, N_1 \wedge N_2]$  and so:

$$V_n = \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)} \left( \frac{k^*}{n} \right) \right).$$

Now, via (A.4),

$$\begin{aligned} V_n - X_n &\leq \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)} \left( \frac{k^*}{n} \right) \right) - \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) \right) \\ &\leq \left( -\frac{1}{2} \widehat{B}_n^{(i_*)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i_*)} \left( \frac{k^*}{n} \right) \right) - \left( -\frac{1}{2} \widehat{B}_n^{(i_*)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i_*)} \left( \frac{1}{2} \right) \right) \\ &= \widehat{B}_n^{(i_*)} \left( \frac{k^*}{n} \right) - \widehat{B}_n^{(i_*)} \left( \frac{1}{2} \right) \end{aligned}$$

$$\begin{aligned}
&\leq \max_{t \in [\frac{1}{2}, \frac{N_{i_*}}{n}]} \left( \widehat{B}_n^{(i_*)}(t) - \widehat{B}_n^{(i_*)} \left( \frac{1}{2} \right) \right) \\
&\leq \bigvee_{i=1,2} \max_{t \in [\frac{1}{2}, \frac{N_i}{n}]} \left( \widehat{B}_n^{(i)}(t) - \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) \right).
\end{aligned}$$

**Inequality (A.6) replacing (15) of [HLM] and its proof.** If  $N_1 \wedge N_2 \geq n/2$ , then  $X_n \leq V_n$  and similarly if the maximum defining  $X_n$  is attained for some  $t \leq (N_1 \wedge N_2)/n$ , then  $X_n = V_n$ . Therefore, the remaining case in comparing  $X_n$  and  $V_n$  consists in  $N_1 \wedge N_2 \leq n/2$  and a maximum defining  $X_n$  attained at some  $t^* \in [(N_1 \wedge N_2)/n, 1/2]$ . In this case,

$$X_n = \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)}(t^*) \right),$$

and

$$V_n \geq \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)} \left( \frac{N_1 \wedge N_2}{n} \right) \right). \quad (\text{A.5})$$

Again, denote by  $i_*$  the index for which the minimum in (A.5) is attained. Then,

$$\begin{aligned}
X_n - V_n &\leq \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)}(t^*) \right) - \bigwedge_{i=1,2} \left( -\frac{1}{2} \widehat{B}_n^{(i)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i)} \left( \frac{N_1 \wedge N_2}{n} \right) \right) \\
&\leq \left( -\frac{1}{2} \widehat{B}_n^{(i_*)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i_*)}(t^*) \right) - \left( -\frac{1}{2} \widehat{B}_n^{(i_*)} \left( \frac{1}{2} \right) + \widehat{B}_n^{(i_*)} \left( \frac{N_1 \wedge N_2}{n} \right) \right) \\
&= \widehat{B}_n^{(i_*)}(t^*) - \widehat{B}_n^{(i_*)} \left( \frac{N_1 \wedge N_2}{n} \right) \\
&\leq \max_{t \in [\frac{N_1 \wedge N_2}{n}, \frac{1}{2}]} \left( \widehat{B}_n^{(i_*)}(t) - \widehat{B}_n^{(i_*)} \left( \frac{N_1 \wedge N_2}{n} \right) \right) \\
&\leq \bigvee_{i=1,2} \max_{t \in [\frac{N_1 \wedge N_2}{n}, \frac{1}{2}]} \left( \widehat{B}_n^{(i)}(t) - \widehat{B}_n^{(i)} \left( \frac{N_1 \wedge N_2}{n} \right) \right). \quad (\text{A.6})
\end{aligned}$$

Since (15) of [HLM] has to be replaced by (A.6), instead of (16) of [HLM], we now have to prove that for  $i = 1, 2$ :

$$\max_{t \in [\frac{N_1 \wedge N_2}{n}, \frac{1}{2}]} \left( \widehat{B}_n^{(i)}(t) - \widehat{B}_n^{(i)} \left( \frac{N_1 \wedge N_2}{n} \right) \right) \xrightarrow{\mathbb{P}} 0. \quad (\text{A.7})$$

The difference with (16) of [HLM] is that  $N$ , therein, is now replaced by  $N_1 \wedge N_2$  which is now more complex since one of the two quantities  $N_1$  or  $N_2$  is not independent of  $\widehat{B}_n$ . To prove (A.7), and so as not to further burden the notation, the superscript  $i$  in the Brownian approximation  $\widehat{B}_n^{(i)}$  is dropped. First, let

$$C_n^1 = \left\{ \left| N_1 - \frac{n}{2} \right| \leq \sqrt{n \ln n} \right\},$$

and, in a similar fashion, define  $C_n^2$  by replacing  $N_1$  with  $N_2$ . Clearly,  $\lim_{n \rightarrow +\infty} \mathbb{P}((C_n^1)^c) = \lim_{n \rightarrow +\infty} \mathbb{P}((C_n^2)^c) = 0$ . Next, for  $\varepsilon > 0$ , let

$$A_n = \left\{ \max_{t \in [\frac{N_1 \wedge N_2}{n}, \frac{1}{2}]} \left| \widehat{B}_n(t) - \widehat{B}_n\left(\frac{N_1 \wedge N_2}{n}\right) \right| \geq \varepsilon \right\}.$$

Then,

$$\mathbb{P}(A_n) \leq \mathbb{P}(A_n \cap C_n^1 \cap C_n^2) + \mathbb{P}((C_n^1)^c) + \mathbb{P}((C_n^2)^c), \quad (\text{A.8})$$

and since on  $C_n^1$  (resp.  $C_n^2$ ),  $N_1 \geq n/2 - \sqrt{n} \ln n$  (resp.  $N_2 \geq n/2 - \sqrt{n} \ln n$ ),

$$\mathbb{P}(A_n \cap C_n^1 \cap C_n^2) \leq \mathbb{P}\left(\left\{ \max_{k=\frac{n}{2}-\sqrt{n} \ln n, \dots, \frac{n}{2}} \left| \sum_{j=N_1 \wedge N_2}^k \xi_j \right| \geq \varepsilon \sqrt{2n} \right\} \cap C_n^1 \cap C_n^2\right), \quad (\text{A.9})$$

where the random variables  $\xi_j$  are iid with mean zero and variance one and assuming that  $n/2 - \sqrt{n} \ln n$  and  $n/2$  are integers (if not replace throughout, the first value by its integer part and the second by its integer part plus one). To deal with (A.9), first note that on  $C_n^1 \cap C_n^2$ ,  $N_1 \wedge N_2 \in [\frac{n}{2} - \sqrt{n} \ln n, n]$ , the right-hand side of (A.9) is clearly upper-bounded by

$$\begin{aligned} & \mathbb{P}\left(\left\{ \max_{\frac{n}{2}-\sqrt{n} \ln n \leq \ell \leq k \leq \frac{n}{2}} \left| \sum_{j=\ell}^k \xi_j \right| \geq \varepsilon \sqrt{\frac{n}{2}} \right\} \cap C_n^1 \cap C_n^2\right) \\ & \leq \mathbb{P}\left(\left\{ \max_{\frac{n}{2}-\sqrt{n} \ln n \leq k \leq \frac{n}{2}} \left| \sum_{j=k}^{n/2} \xi_j \right| \geq \frac{\varepsilon}{2} \sqrt{\frac{n}{2}} \right\} \cap C_n^1 \cap C_n^2\right) \end{aligned} \quad (\text{A.10})$$

$$\leq \frac{8 \ln n}{\varepsilon^2 \sqrt{n}}, \quad (\text{A.11})$$

where the inequality in (A.10) follows from the bound

$$\begin{aligned} \max_{\frac{n}{2}-\sqrt{n} \ln n \leq \ell \leq k \leq \frac{n}{2}} \left| \sum_{j=\ell}^k \xi_j \right| & \leq \max_{\frac{n}{2}-\sqrt{n} \ln n \leq \ell \leq k \leq \frac{n}{2}} \left( \left| \sum_{j=k}^{n/2} \xi_j \right| + \left| \sum_{j=\ell}^{n/2} \xi_j \right| \right) \\ & \leq 2 \max_{\frac{n}{2}-\sqrt{n} \ln n \leq k \leq \frac{n}{2}} \left| \sum_{j=k}^{n/2} \xi_j \right|, \end{aligned}$$

while the one in (A.11) is Kolmogorov's maximal inequality. Therefore, the right-hand side of (A.9) converges to zero, finishing, via (A.8), the proof of (A.7).  $\square$

# Acknowledgments

This research is supported in part by the grant #246283 from the Simons Foundation and by a Simons Foundation Fellowship grant #267336. The first author thanks the School of Mathematics of the Georgia Institute of Technology for several visits during which part of this work was done. The second author thanks the Centre Henri Lebesgue of the Université de Rennes 1, the Département MAS of École Centrale Paris, the LPMA of the Université Pierre et Marie Curie and CIMAT, Gto, Mexico for their hospitality, while this work was in progress. Both author thank an anonymous referee for valuable comments which helped to improve this manuscript.

# References

- [BDJ] J. Baik, P. Deift, and K. Johansson. *On the distribution of the length of the longest increasing subsequence of random permutations*. J. Amer. Math. Soc., 12(4), pp.1119–1178, 1999.
- [Bar] Y. Baryshnikov. *GUEs and queues*, Probab. Theory Relat. Fields vol. 119, pp. 256–274, 2001.
- [BGH] F. Benaych-Georges, C. Houdré. *GUE minors, maximal Brownian functionals and longest increasing subsequences in random words*. Markov Processes Relat. Fields, vol. 21, pp. 109–126, 2015.
- [Bil] P. Billingsley. *Convergence of probability measures*, Wiley series in Probability and Statistics, 2nd Edition, 1999.
- [BM] T. Bodineau, J. Martin. *A universality property for last-passage percolation paths close to the axis*. Elec. Comm. Prob. vol. 10, pp. 105–112, 2005
- [BH] J.-C. Breton, C. Houdré. *Simultaneous asymptotics for the shape of random Young tableaux with growingly reshuffled alphabets*. Bernoulli, vol. 16, no. 2, pp. 471–492, 2010.
- [CZFYZ] W.T. Chan, Y. Zhang, S.P.Y. Fung, D. Ye and H. Zhu. *Efficient algorithms for finding a longest common increasing subsequence*. Lecture Notes in Comput. Sci., vol. 3827, Springer, Berlin, pp. 655–674, 2005.
- [DKFPWS] A.L. Delcher, S. Kasif, R.D. Fleischmann, J. Peterson, O. White and S.L. Salzberg. *Alignment of whole genomes*. Nucleic Acids Research, vol. 27, no. 11, pp. 2369–2376, 1999.
- [GW] P. W. Glynn, W. Whitt. *Departure from many queues in series*. Ann. Appl. Probab., 1(4), pp. 546–572, 1991.
- [GTW] J. Gravner, C. A. Tracy, H. Widom. *Limit theorems for height fluctuations in a class of discrete space and time growth models*, J. Stat. Phys. vol. 102, pp. 1085–1132, 2001.
- [HI] C. Houdré and Ü. Islak. *A central limit theorem for the length of the longest common subsequences in random words*. ArXiv #Math.PR/1408.1559v3, 2015.
- [HLM] C. Houdré, J. Lember, H. Maztinger. *On the longest common increasing binary subsequence*. C.R. Acad. Sci., Paris Ser. I, vol. 343, pp. 589–594, 2006.
- [HL] C. Houdré, T. Litherland. *On the longest increasing subsequence for finite and countable alphabets*, in High Dimensional Probability V: The Luminy Volume (Beachwood, Ohio, USA: Institute of Mathematical Statistics), pp. 185–212, 2009.
- [HX] C. Houdré, H. Xu. *On the limiting shape of Young diagrams associated with inhomogeneous random words*, in: High Dimensional Probability VI: The Banff volume Progress in Probability, 66, Birkhauser, pp. 277–302, 2013.
- [ITW1] A. Its, C. A. Tracy, H. Widom. *Random words, Toeplitz determinants, and integrable systems. I*. Random matrix models and their applications, pp. 245–258, Math. Sci. Res. Inst. Publ., vol. 40, Cambridge Univ. Press, Cambridge, 2001.
- [ITW2] A. Its, C. A. Tracy, H. Widom. *Random words, Toeplitz determinants, and integrable systems. II*. Advances in nonlinear mathematics and science. Phys. D., vol. 152–153, pp. 199–224, 2001.

- [Joh] K. Johansson. *Discrete orthogonal polynomial ensembles and the Plancherel measure*. Ann. of Math. (2) 153 (2001), no. 1, 259–296.
- [Ker] S. Kerov. *Asymptotic Representation Theory of the Symmetric Group and its Applications in Analysis*, Vol. 219. AMS, Translations of Mathematical Monographs, 2003. (Russian edition: D. Sci thesis, 1994)
- [Rom] D. Romik. *The surprising mathematics of longest increasing subsequences*. Cambridge University Press, 2014.
- [Sak] Y. Sakai. *A linear space algorithm for computing a longest common increasing subsequence*. Information Processing Letters, vol. 99, pp. 203–207, 2006.
- [Sep] T. Seppäläinen. *A scaling limit for queues in series*. Ann. Appl. Probab., 7(4), pp. 855–872, 1997.
- [TW] C. A. Tracy, H. Widom. *On the distribution of the lengths of the longest increasing monotone subsequences in random words*. Probab. Theor. Rel. Fields. vol. 119, pp. 350–380, 2001.
- [VK] A. M. Vershik, S. V. Kerov. *Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux*. Soviet Math. Dokl. (English translation), 233(1–6): pp. 527–531, 1977.